DOCUMENT RESUME

ED 481 836 TM 035 348

AUTHOR Pedulla, Joseph J.; Abrams, Lisa M.; Madaus, George F.; Russell, Michael K.; Ramos, Miguel A.; Miao, Jing

TITLE Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers.

INSTITUTION National Board on Educational Testing and Public Policy, Chestnut Hill, MA.

PUB DATE 2003-03-00

NOTE 151p.

PUB TYPE Reports - Research (143)

EDRS PRICE EDRS Price MF01/PC07 Plus Postage.

DESCRIPTORS Elementary Secondary Education; *High Stakes Tests; National Surveys; *State Programs; *Teacher Attitudes; Teacher Surveys; *Teachers; *Test Use; Testing Programs

IDENTIFIERS *Testing Effects

ABSTRACT

Results from a national survey of teachers are reported for five types of state testing programs, those with: (1) high stakes for districts, schools, or teachers, and students; (2) high stakes for districts, schools, and teachers, and moderate stakes for students; (3) high stakes for districts, schools, and teachers, and low stakes for students; (4) moderate stakes for districts, schools, and teachers, and high stakes for students; and (5) moderate stakes for districts, schools, and teachers, and low stakes for students. Of the 12,000 teachers who received surveys, 4,195 returned responses. At least two themes emerged from these survey data. In several areas, teachers' responses differ significantly when analyzed by the severity of the stakes attached to test results. Pressure on teachers, emphasis on test preparation, time devoted to test content, and views on accountability are such areas. The second theme is that views of elementary, middle, and high school teachers regarding the effects of their state's test differed from each other in areas such as school climate and classroom use of test results. There are also instances in which stakes and grade level combined show interesting patterns in teachers' responses, and areas in which there are no differences. The summary is organized by major areas surveyed, and within each area, findings are presented for stakes levels, grade levels, and stakes combined with grade levels. Five appendixes contain supplemental information and data tables. (Contains 16 figures, 88 tables, and 53 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 481 836

Boston College

Perceived Effects of State-Mandated Testing Programs on Teaching and Learning:

Findings from a National Survey of Teachers

National Board on Educational Testing and Public Policy

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J. Pedulla
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Joseph J. Pedulla, Lisa M. Abrams, George F. Madaus,
Michael K. Russell, Miguel A. Ramos, and Jing Miao

Lynch School of Education

Boston College

March 2003

NBETPP

ACKNOWLEDGEMENTS

We would like to thank The Atlantic Philanthropies Foundation for generously funding this research. We are grateful as well for the sound advice of our advisory board: Albert Beaton, Robert Hauser, Henry Levin, Audrey Qualls, and Dan Stufflebeam. We would also like to recognize the efforts of Marguerite Clarke, Catherine Horn, and Jie Li who assisted with the development, administration, and analysis of the survey. In addition, we want to thank Irwin Blumer, Dana Diaconu, Amie Goldberg, Tom Hoffmann, Arnold Shore, and Stuart Yeh for their contributions to this study. Above all, we are tremendously grateful to the thousands of teachers nationwide who took time out of their very busy school day to complete our survey. The conclusions and opinions presented here are those of the authors and do not necessarily reflect the views of The Atlantic Philanthropies Foundation; we are solely responsible for any errors.

IN MEMORIAM

Audrey Qualls

In early January 2003 we were saddened to learn of the death of our advisory board member, Audrey Qualls. Audrey was Professor of Measurement and Statistics at The University of Iowa. Throughout her career, she made valuable contributions to the fields of testing and measurement. She also made great contributions to this project. Her humor and perceptive remarks as well as her willingness to review and provide feedback on project materials helped guide us through much of our early work. Audrey was a wonderful scholar, intellect, and friend. She will be missed.

CONTENTS

I. List of Tables iv

II. List of Figures viii

III. Executive Summary 1

IV. Introduction 10

V. Background 11

VI. Methodology 16

VII. Results of the National Survey 22

  1. School Climate 22
  2. Pressure on Teachers 28
  3. Alignment of Classroom Practices with the State Test 34
  4. Perceived Value of the State Test 39
  5. Impact of the State Test on Content and Modes of Instruction 47
  6. Test Preparation and Administration 56
  7. Unintended Consequences of the State Test 77
  8. Use of Test Results 83

VIII. Summary and Conclusions 112

IX. End Notes 123

X. References 124

XI. Appendices 127

LIST OF TABLES

Table 1 Basic Sampling Frame 18

Table 2 Means on the School Climate Scale by School Type 23

Table 3 Views on School Climate: Percent Agreement by Stakes Level 25

Table 4 Views on School Climate: Percent Agreement by School Type 27

Table 5 Means on the Pressure Scale by Stakes Level and School Type. 29

Table 6 Pressure on Teachers: Percent Agreement by Stakes Level 31

Table 7 Pressure on Teachers: Percent Agreement by School Type 33

Table 8 Means on the Alignment Scale by Stakes Level 35

Table 9 Alignment with the State Test: Percent of Agreement by Stakes Level 37

Table 10 Alignment with the State Test: Percent of Agreement by School Type 38

Table 11 Value of State Test: Percent of Agreement by Stakes Level 41

Table 12 Test as a Measure of Achievement: Percent Agreement by Stakes Level 42

Table 13 Media Coverage: Percent Agreement by Stakes Level 43

Table 14 Value of State Test: Percent of Agreement by School Type 44

Table 15 Test as a Measure of Achievement: Percent Agreement by School Type 45

Table 16 Media Coverage: Percent Agreement by School Type 46

Table 17 Items Comprised by the Tested Areas, Non-Core Content, and Classroom Activities Scales 48

Table 18 Means on the Tested Areas, Non-Core Content, and Classroom Activities Scales by Stakes Level 49

Table 19 Means on the Tested Areas, Non-Core Content, and Classroom Activities Scales by School Type 49

Table 20 Tested and Non-tested Content Areas: Percent Reporting Change in Instructional Time 50

Table 21 Non-Core Content Areas: Percent Reporting Change in Instructional Time 51

Table 22 Classroom Activities: Percent Reporting Change in Instructional Time 53

Table 23 Methods of Instruction: Percent Agreement by Stakes Level 54

Table 24 Test Preparation Strategies: Percent Reporting by Stakes Level 57

Table 25 Test Preparation Strategies: Percent Reporting by Stakes Level and School Type....................59

Table 26 Test Preparation Time: Percent Reporting by Stakes Level 60

Table 27 Test Preparation Time: Percent Reporting by Stakes Level and School Type 61

Table 28 Timing of Test Preparation: Percent Reporting by Stakes Level....................................62

Table 29 Timing of Test Preparation: Percent Reporting by Stakes Level and School Type 64

Table 30 Content of Test Preparation Material: Percent Reporting by Stakes Level 65

Table 31 Content of Test Preparation Material: Percent Reporting by
Stakes Level and School Type 66

Table 32 Groups Targeted for Test Preparation: Percent Reporting by Stakes Level 67

Table 33 Groups Targeted for Test Preparation: Percent Reporting by
Stakes Level and School Type 69

Table 34 Teachers' Test Administration Practices: Percent Reporting by Stakes Level 70

Table 35 Teachers' Test Administration Practices: Percent Reporting by
Stakes Level and School Type 71

Table 36 Schoolwide Motivational Strategies: Percent Reporting by Stakes Level 72

Table 37 Schoolwide Motivational Strategies: Percent Reporting by
Stakes Level and School Type 75

Table 38 Test Leading to Grade Retention: Percent Agreement by
Stakes Level and School Type 79

Table 39 Test Leading to Dropping Out: Percent Agreement by Stakes Level and School Type 80

Table 40 Computer Use Precluded by Test Format: Percent Agreement by
Stakes Level and School Type 81

Table 41 Policy Ban on Computer Use in Writing Instruction:
Percent Agreement by Stakes Level and School Type....................................82

Table 42 Items Comprised by School, Student, and Teacher/Administrator
Accountability Scales 84

Table 43 Means on the Accountability Scales by Stakes Level 85

Table 44 Use of Test Results for School Accountability: Percent Reporting by Stakes Level 87

Table 45 Use of Test Results for School Accountability:
Percent Reporting by Stakes Level and School Type 87

Table 46 Use of Test Results to Promote or Retain Students:
Percent Reporting by Stakes Level 89

Table 47 Use of Test Results to Promote or Retain Students:
Percent Reporting by Stakes Level and School Type 89

Table 48 Use of Test Results to Evaluate Teachers/Administrators:
Percent Reporting by Stakes Level 91

Table 49 Use of Test Results to Evaluate Teachers/Administrators:
Percent Reporting by Stakes Level and School Type 91

Table 50 District-Level Use of Test Results: Percent Reporting by Stakes Level 93

Table 51 District-Level Use of Test Results: Percent Reporting by School Type 96

Table 52 Influence of School's Test Results on Teaching: Percent Reporting by Stakes Level 98

Table 53 Influence of Students' Test Results on Teaching: Percent Reporting by Stakes Level 98

Table 54 Influence of School's Test Results on Teaching: Percent Reporting by School Type 99

Table 55 Influence of Students' Test Results on Teaching: Percent Reporting by School Type 100

Table 56 Classroom Use of Test Results: Percent Reporting by Stakes Level 101

Table 57 Classroom Use of Test Results: Percent Reporting by School Type 103

Table 58 Characteristics of the Individual Student Reports: Percent Reporting by Stakes Level 105

Table 59 Characteristics of the School Reports: Percent Reporting by Stakes Level 105

Table 60 Characteristics of the District Reports: Percent Reporting by Stakes Level 106

Table 61 Characteristics of the Individual Student Reports: Percent Reporting by School Type 106

Table 62 Characteristics of the School Reports: Percent Reporting by School Type 107

Table 63 Characteristics of the District Reports: Percent Reporting by School Type 107

Table 64 Adequacy of Professional Development: Percent of Teachers Reporting 109

Table 65 Adequacy of Professional Development: Percent of H/H and M/L Teachers Reporting 109

Table 66 Adequacy of Professional Development: Percent Reporting by School Type 111

Table B1 State Testing Program Classifications....135

Table C1 Sampling Stratification by School Type 136

Table C2 Sampling Stratification by School Type and Subject Area 137

Table C3 Final Sampling Frame....137

Table D1 Characteristics of Survey Respondents 138

Table E1 School Climate Scale Summary 139

Table E2 ANOVA Results for Stakes Level and School Type on the School-Climate Scale 139

Table E3 Pressure Scale Summary 140

Table E4 ANOVA Results for Stakes Level and School Type on the Pressure Scale 140

Table E5 Alignment Scale Summary 141

Table E6 ANOVA Results for Stakes Level and School Type on the Alignment Scale....141

Table E7 Perceived-Value Scale Summary 142

Table E8 ANOVA Results for Stakes Level and School Type on the Perceived Value Scale 143

Table E9 Tested-Areas, Non-Core Content, Classroom Activities Scales Summary 144

Table E10 ANOVA Results by Stakes Level and School Type for Tested Areas Scale 145

Table E11 ANOVA Results by Stakes Level and School Type for Non-Core Content Scale....145

Table E12 ANOVA Results by Stakes Level and School Type for Classroom Activities Scale....145

Table E13 School, Student, Teacher/Administrator Accountability Scales Summary 146

Table E14 ANOVA Results for Stakes Level and School Type on the School Accountability Scale 147

Table E15 ANOVA Results for Stakes Level and School Type on the Student Accountability Scale 147

Table E16 ANOVA Results for Stakes Level and School Type on the Teacher/Administrator Accountability Scale 147

LIST OF FIGURES

Figure 1 Main Effects in ANOVA 21

Figure 2 Main Effects and Interaction Effects in ANOVA 21

Figure 3 School Climate: Agreement of Elementary and Middle School Teachers vs. High School Teachers 24

Figure 4 Pressure Scale Means: School Type by Stakes Level 30

Figure 5 Alignment with the State Test: Agreement by H/H, H/M, H/L and M/H vs. M/L Stakes States 36

Figure 6 Use of Test Preparation Strategies: H/H and M/H vs. M/L Stakes States 58

Figure 7 Test Preparation Hours: H/H and M/H vs. M/L Stakes States 60

Figure 8 Test Preparation Timing: H/H and M/H vs. M/L Stakes States 63

Figure 9 Test Preparation Content: H/H and M/H vs. M/L Stakes States 65

Figure 10 Target of Test Preparation: H/H and M/H vs. M/L Stakes States 68

Figure 11 Unethical Test Administration Practices: H/H and M/H vs. M/L Stakes States 70

Figure 12 Use of Schoolwide Motivational Strategies: H/H and M/H vs. M/L Stakes States 73

Figure 13 Agreement for Unintended Consequences: H/H, H/M, H/L, M/H vs. M/L Stakes States 78

Figure 14 Appropriateness of Using Test Results for School Accountability: H/H, H/M, H/L and M/H vs. M/L Teachers' Responses 88

Figure 15 Appropriateness of Using Test Results for Student Accountability: H/H, H/M, H/L and M/H vs. M/L Teachers' Responses 90

Figure 16 Appropriateness of Using Test Results for Teacher Accountability: H/H, H/M, H/L and M/H vs. M/L Teachers' Responses 92

EXECUTIVE SUMMARY

Tests have consistently been viewed as a lever to change classroom practices and improve general education. The current emphasis on high-stakes testing resulting from standards-based reform efforts is largely an extension of three decades of testing, with a new emphasis on higher standards and greater academic achievement. In large part, current state tests were designed to serve two functions: to measure student achievement of the state's content standards and to indicate school effectiveness.

To that end, consequences in the form of rewards and sanctions have been attached to test results in an effort to improve teachers' and students' performance. These rewards and sanctions vary from high to low in severity. Generally, they are applied at both the institutional level (districts, schools, administrators, teachers) and the student level — sometimes with similar stakes and sometimes with different stakes. Of particular interest in this study was the relationship between the two levels of accountability (stakes for districts, schools, and/or teachers, and stakes for students) and the effect of state testing programs on classroom practices as witnessed by those who experience their impact firsthand, namely classroom teachers. Consequently, results from the national survey of teachers are reported for five types of state testing programs, those with (1) high stakes for districts, schools, and/or teachers *and* high stakes for students (H/H), (2) high stakes for districts, schools, and/or teachers *and* moderate stakes for students (H/M), (3) high stakes for districts, schools, and/or teachers *and* low stakes for students (H/L), (4) moderate stakes for districts, schools, and/or teachers *and* high stakes for students (M/H), and (5) moderate stakes for districts, schools, and teachers *and* low stakes for students (M/L).

At least two themes emerge from these survey data. First, in several areas teachers' responses differ significantly when analyzed by the severity of the stakes attached to test results. Pressure on teachers, emphasis on test preparation, time devoted to tested content, and views on accountability are such areas. The second theme is that the views of elementary, middle, and high school teachers regarding the effects of their state's test differed from each other in areas such as school climate and classroom use of test results. And then, there are instances when stakes and grade level combined show interesting patterns in teachers' responses; in others there are no differences at all.

This summary is organized like the Findings section, by major areas surveyed. These areas include (1) school climate, (2) pressure on teachers, (3) perceived value of the state test, (4) alignment of classroom practices with the state test, (5) impact on the content and mode of instruction, (6) test preparation and administration, (7) perceived unintended consequences, and (8) accountability and use of test results. Within each area, we present findings for stakes levels, grade levels, and stakes combined with grade levels.

I. School Climate

Items related to school climate dealt with teacher expectations for students, student morale, how conducive the climate was to learning, student motivation, and testing pressure on students. Teachers from high-stakes states were more likely than were teachers from M/L states to report that students felt intense pressure to perform well and were extremely anxious about taking the state test. In states with high stakes for students, three-quarters or more of teachers reported this degree of pressure. This compares with about half of the teachers in low-stakes states. Test-related anxiety and pressure did not negatively influence teachers' expectations of student performance or perceptions of school climate. In states where stakes are high for students, large majorities of teachers (8 in 10) reported that most of their students tried their best on the state test. Although most teachers (7 in 10) indicated that student morale was high, teachers in low-stakes states were more likely to report this than were their colleagues in high-stakes states.

Elementary and middle school teachers were more positive about school climate than were their high school counterparts. Nonetheless, more elementary and middle school teachers than high school teachers reported that their students are extremely anxious and are under intense pressure because of the state test. In other words, the psychological impact was perceived to be greater at the elementary level, yet this did not seem to negatively affect the general atmosphere of the school.

II. Pressure on Teachers

Items related to pressure on teachers dealt with pressure from administrators and parents to improve test scores, pressure to limit teaching to what is tested and to change teaching methods in ways that are not beneficial, and teachers' discontent with their profession (low morale or wanting to transfer out of tested grades). In general, teachers in high-stakes states reported feeling more pressure than those in lower-stakes states. However, regardless of the consequences attached to the state test, teachers reported similar feelings of pressure from parents to raise test scores and similar views on school morale. A large majority of teachers felt that there is so much pressure for high scores on the state-mandated test that they have little time to teach anything not covered on the test. This view was most pronounced in states where high levels of accountability are demanded of districts, schools, teachers, and students. This finding supports the contention that state testing programs have the effect of narrowing the curriculum. Also, teachers in high-stakes states were more likely than those in low-stakes states to report that they feel pressure from the district superintendent, and to a lesser degree from their building principal, to raise test scores. While most teachers reported such pressure, it was significantly lower for those in low-stakes than in high-stakes states. Between 3 in 10 and 4 in 10 teachers in high-stakes states compared with 2 in 10 of their counterparts in low-stakes states reported that teachers at their school want to transfer out of the tested grades.

Generally, elementary teachers reported feeling more pressure than high school teachers, while middle school teachers were somewhere in between. Further, elementary and middle school teachers in states with high stakes for districts, schools, teachers, and students reported the greatest feelings of test-related pressure as compared with their counterparts in other testing programs. A substantial majority of teachers at each grade level indicated that state testing programs have led them to teach in ways that contradict their ideas of sound instructional practices; this view was particularly pronounced among elementary teachers. This finding is a particularly distressing one and highlights the fact that state testing programs can have unintended negative effects.

III. Alignment of Classroom Practices with the State Test

Items related to alignment of classroom practices with the state test dealt with compatibility between the test and the curriculum, instruction, texts, and teacher-made tests. Teachers in the H/H and H/L groups indicated greater alignment at the scale score level than did teachers in the other groups. At the individual item level, teachers in low-stakes states more often than teachers in high-stakes states found that teaching the state standards resulted in better test performance. Far more teachers in high-stakes states said their own tests reflected the format of the state test than did teachers in low-stakes states. A similar pattern occurred with regard to the content of teacher-made tests, although the differences were not as large.

Elementary teachers held the most positive opinion of state curricular standards but were less positive than high school teachers about the compatibility of their instructional texts and materials with the state tests. This may be due to the fact that unlike high school teachers, who generally teach one subject, elementary teachers have to deal with several tested subjects per grade. With far more texts and materials, there is more room for disparity. A majority of all teachers were positive in their opinions of their state's curricular standards, and the vast majority indicated that their district's curriculum was aligned with the state test.

IV. Perceived Value of the State Test

Items related to the perceived value of the state test dealt with the accuracy of inferences that can be made from the test about quality of instruction, student learning, school effectiveness, and differences among various groups; the adequacy and appropriateness of media coverage of test results; and the cost/benefit ratio of the testing program. Teachers in high-stakes states, more so than those in low-stakes states, reported that the test brought much-needed attention to education issues. It should be noted that it was a minority of teachers across all stakes levels who agreed with this assessment of the power of the state test to call public attention to educational issues.

Elementary teachers felt to a greater degree than either middle or high school teachers that the state test measured achievement of high standards. Middle school teachers more often agreed with this item than did high school teachers. More elementary teachers thought that the test did not accurately measure what minority students know than did middle or high school teachers. Both elementary and middle school teachers felt to a greater degree than high school teachers that the test score differences from year to year reflected changes in the characteristics of students rather than changes in school effectiveness. Elementary teachers, more than middle or high school teachers, indicated that media reporting about the state test was not accurate.

About three-quarters of all teachers, regardless of stakes or grade level, found that the benefits of the testing program were not worth the time and money involved. A similar proportion felt that the media coverage of state-mandated testing issues was unfair to teachers and inaccurately portrayed the quality of education and the complexity of teaching. Across all stakes levels, 9 in 10 teachers did not regard the state test as an accurate measure of what ESL students know and can do, and 4 in 10 teachers reported that teachers in their school could raise test scores without improving learning.

V. Impact on the Content and Mode of Instruction

Items regarding the impact on classroom instruction dealt with changes in the amount of time spent on a variety of activities and with the influence of the testing program on pedagogical practices and instructional emphasis. The items clustered into 3 scales: (1) impact on tested subject areas, (2) impact on non-core subject areas, and (3) impact on student and class activities.

More teachers in states with high stakes for students than in states with lesser stakes indicated that they spent more time on instruction in tested areas and less on instruction in non-core subject areas (e.g. fine arts, physical education, foreign languages, industrial/vocational education) and on other activities (e.g. field trips, enrichment activities). In general, the influence of state testing programs on teachers' instructional practices is more closely related to the stakes for students than those for schools.

More elementary and middle school teachers than high school teachers reported that they increased the amount of time spent on tested areas and decreased the time spent on non-core subject areas and on other activities. The impact of testing programs is generally stronger in elementary and middle schools than in high schools.

Across all types of testing programs, teachers reported increased time spent on subject areas that are tested and less time on areas not tested. They also reported that testing has influenced the time spent using a variety of instructional methods such as whole-group instruction, individual-seat work, cooperative learning, and using problems similar to those on the test.

VI. Test Preparation

Teachers responded to a series of items related to preparing their students for the state-mandated test (e.g. on test preparation methods used and amount of time spent on test preparation). Teachers in states with high-stakes tests are much more apt than their counterparts in states with lower-stakes tests to engage in test preparation earlier in the school year; spend more time on such initiatives; target special groups of students for more intense preparation; use materials that closely resemble the test; use commercially or state-developed test-specific preparation materials; use released items from the state test; and try to motivate their students to do well on the state test.

Teachers in high-stakes states were more likely to report that they focused test preparation on students who were on the border either of passing or of moving to the next performance level. Elementary teachers in high-stakes states reported spending more time on test preparation than did their high school counterparts. Further, elementary teachers were more apt to report engaging in test preparation throughout the year than were middle or high school teachers.

Elementary teachers in states with high stakes for schools and students were twice as likely as teachers in the low-stakes states to report that their test preparation content was very similar to the content of the state test. When asked whether summer school should be required or recommended as a motivator roughly half of elementary and middle school teachers and a third of secondary teachers in the H/H states responded affirmatively compared with fewer than 1 in 10 teachers across all grade levels in the low-stakes states. Retention in grade as a motivator was selected by a quarter of elementary teachers, a third of middle school teachers, and 1 in 5 high school teachers in H/H states, while the frequency in the M/L states never reached 5% at any grade level.

VII. Unintended Consequences of the State Test

Survey items in this area dealt with the effect of state testing programs on the instructional use of technology — specifically the use of computers in writing instruction and the effect of the state test on decisions related to persistence, including decisions about grade retention and dropping out of high school. One-third of teachers in H/H states compared with one-fifth of those in M/L states said their school does not use computers when teaching writing because the state test is handwritten. Roughly one-fourth of teachers in states with high stakes for both schools and students, and one-tenth in the other high-stakes states, agreed that the test has caused retention in grades, contrasted with only 3% of teachers in low-stakes states. As for dropouts, 25% of teachers in states with high stakes for students compared with 10% of all other teachers state that the testing caused many students to drop out of high school.

A majority of teachers across stakes and grade levels disagreed with all of the four unintended consequences described in this section – teachers not using computers to teach writing because the state writing test is handwritten, the district forbidding the use of computers in writing instruction, the test causing many students to drop out of high school, and the test having caused many students to be retained in grade.

VIII. Use of StateTest Results

Teachers' views on the use of the state test results fell into the following four categories: (1) district-level use, (2) classroom-level use, (3) the reporting of test results, and (4) professional development and resources. Results for each area will be presented in turn.

A. Views on District-Level Use

Items in this area dealt with the use of state test results for three accountability purposes: school, student, and teacher/administrator accountability. Teachers in H/H states viewed the use of state tests for school, student, and teacher/administrator accountability as slightly less inappropriate than did teachers in other states. Further, student accountability was the most appropriate of the three uses (between moderately appropriate and moderately inappropriate, a neutral view), and teacher/administrator accountability the least appropriate. Although teachers in H/H states viewed the use of test results for accountability somewhat more favorably (or at least less unfavorably) than their counterparts in other states, their opinions were still at the neutral to unfavorable end of the spectrum relative to teachers in states where the stakes are not as high. This less unfavorable view could be a result of teachers' being more comfortable with test use for accountability, or simply being resigned to such uses. Many more teachers in H/H states (25%) said that their students' test results influence their teaching on a daily basis than did teachers in the states with lower stakes (10%).

Greater percentages of high school than elementary or middle school teachers, not surprisingly, reported that test results were used in their district to make decisions about graduation. Generally, awareness of how test results are used was lower at the high school level than in elementary or middle schools. This finding is reasonable for decisions about placement in groups by ability or in special education, which are generally made before high school and are simply carried forward independently of state test results. It makes less sense, however, for other uses (e.g. ranking schools publicly or holding schools accountable), where district-level use should be the same across all three school types.

Teachers, on average across all the states, were neutral regarding the use of state test results for student accountability. Their use for school accountability was seen on average as moderately inappropriate, and for teacher/administrator accountability as moderately to very inappropriate. When asked how state tests were actually used in their districts, all teachers most frequently cited use for accountability of schools and districts, ranking schools, and remediating students. Most uses of test results were cited by less than 30% of all teachers and many by less than 10%.

B. Views on Classroom-Level Use

Items in this area dealt with the influence of school- and student-level test results on teaching. Teachers were asked how often school-level and student-level results on the state test affected their teaching. Significantly more teachers (40%) in states with high stakes for schools and students than in low-stakes states (10%) reported that their school's results influenced their teaching on a daily basis. Conversely, a greater percentage of teachers in low-stakes states (25%) indicated that the school's results influenced their teaching a few times a year than teachers in states with high stakes for schools and students (roughly 10%).

Teachers in H/H states tend to use state-mandated test results for classroom decisions to a greater extent than do teachers in low-stakes situations. Teachers in states with high stakes for schools and students used the results the most of any group to plan instruction (60%) and to select instructional materials (50%); teachers in low-stakes states used them the least (40% and 30% respectively). Teachers in states with high stakes for schools and students reported using the results significantly more frequently to give feedback to students than did their counterparts in low-stakes situations. Teachers in H/H states also reported using the results more often than other teachers to evaluate student progress; to group students within the class; and to determine student grades. It should be noted that the latter two uses were chosen by a small percentage of all teachers regardless of stakes level.

State-mandated test results influenced elementary teachers' instruction with much greater frequency than was the case for high school teachers. This may occur because the tests now focus elementary instruction on the standards tested, giving elementary teachers who must teach a variety of subjects much greater direction on what should be taught. These findings may also indicate that the state-mandated tests narrow or shape elementary curriculum to a greater degree than is the case at the high school level. Conversely, high school teachers' instruction may be least influenced by the state tests, because these teachers have always taught a specific subject area (e.g. math or history), and the test is measuring, for the most part, content they were already teaching. Middle school teachers fall somewhere between elementary and high school teachers in terms of subject matter specialization, and therefore the influence of the state test results on their instruction is somewhere between that for the other two groups, although generally closer to the elementary teachers. More elementary teachers reported using the results of the state-mandated test to aid in decisions about instruction, assess their own teaching effectiveness, provide feedback to parents, evaluate students, and group students in their class than did high school teachers. In general, high school teachers are least likely to use state-mandated test results.

Clearly, the stakes attached to the results of the state-mandated tests affect the extent to which teachers use them for various instructional and feedback activities. When the stakes are high for students and teachers, teachers use the results to the greatest extent; when they are low, they tend to use them less often. For 7 of the 8 activities listed, fewer than half of the teachers – regardless of stakes level – indicated that they use the test results to inform their practice, the lone exception being that a majority of all teachers reported using results to plan instruction. Further, very small proportions (less than 10% overall) use the results for student-specific decisions (i.e. grouping students within the class or determining student grades).

C. Views on the Reporting of Test Results

Items in this section dealt with the various reports on test results that teachers receive: individual student reports, and school- and district-level reports. A majority of all teachers either agreed or strongly agreed that the individual student reports and the school and district reports are easy to interpret and provide useful information. Significantly more teachers (though still only 10%) in the states with low stakes were unfamiliar with the school and district reports than were teachers in any of the three high-stakes groups. High school teachers were the least familiar with the various reports. Between 10% and 20% reported that they have never seen these reports. Significantly fewer high school teachers than elementary or middle school teachers agreed that the reports provide useful information. Elementary teachers were the most familiar with the school reports; less than 10% reported that they had never seen them.

D. Professional Development and Resource Personnel

Items in this section dealt with the adequacy of professional development related to the state testing program and the availability of someone in the school to deal with and answer questions about the program. The vast majority of all teachers (80%) indicated that they do have someone to turn to at their school to obtain accurate information about the state-mandated testing program. The sole difference occurred between teachers in states with high stakes for students and schools and those in states with low stakes (80% vs. 70%). More teachers in states where the stakes are high viewed the professional development as adequate than did teachers where the stakes are low. Conversely, greater proportions of teachers in low-stakes situations indicated that there is no professional development related to test preparation, interpretation, and use of test results. A significantly smaller percentage of high school teachers also indicated that the professional development activities focused on test preparation, interpretation, and use of test results are less adequate or nonexistent than did elementary or middle school teachers. The majority of all teachers viewed the professional development related to areas concerning implementation of the state-mandated testing program as adequate.

Conclusions

This study shows that the severity of consequences attached to state tests affects the instruction students receive. Generally, as the stakes increase, so does the influence of the test; and in some cases, this influence varies for elementary, middle, and high school teachers within the same testing program. Further, the combination of stakes and grade levels produced significant differences, generally indicating that instruction at the lower grades in high-stakes states is most affected by the state test. However, in some areas there were no differences among stakes and grade levels; these findings were also of interest.

For the most part, the views of teachers in states with high stakes for both students and teachers (or schools and districts), i.e. H/H states, about the effect of state testing programs differed from those of teachers in states where the stakes were low (M/L states). The differences were in the expected direction: teachers in high-stakes situations, particularly in H/H states, reported feeling more pressure to have their students do well on the test, to align their instruction with the test, to engage in more test preparation, and so forth. In many instances, results from teachers in states where the stakes were low for students but high for schools (H/L) were very similar to those for teachers in H/H states.

Elementary teachers often indicated that they are most affected by the statewide testing program. For example, they reported more time spent on instruction in tested areas, less time spent on instruction in non-tested areas, more time spent on test preparation, and greater impact on their instructional practices than did secondary teachers.

The findings in this report need to be examined by policymakers and educators in their own state to determine whether the effects of the state test, as reported here by teachers, are the desired ones. To the extent that undesired effects are occurring, the testing program should be modified to minimize them. Only by listening to what teachers tell us is happening as a result of these testing programs can we be confident that these programs are having the intended effect. Teachers are on the front line every day. Their voice on this issue must be heard; their opinions must enter into the formation of sound testing policy. While some states do involve teachers in the formulation of the testing program, others do not. Even in states that do so, the number of teachers involved is small. We hope the findings presented here give voice to a broader cross-section of teachers than has heretofore been available on issues related to statewide testing programs, and that they spur more teacher input in the future.

INTRODUCTION

During the last decade every state except Iowa has adopted state curriculum frameworks or content standards. In addition, all states with the exception of Nebraska have implemented an assessment program designed to measure student achievement of these curricular standards (Quality Counts, 2002). By 2008, almost half of the states (24) will require students to pass a state test in order to graduate; this requirement will affect 70% of students nationwide (Center on Education Policy, 2002). High-stakes testing policies have a far-reaching impact on the education of students and consequently on their future academic and employment opportunities.

Education reform efforts since 1983 have generally had three main components: (1) educational goals or standards, (2) a test designed to measure the degree to which these goals have been achieved, and (3) high stakes attached to the results, which are intended to influence the behavior of teachers and students. Many believe that the high-stakes component of state testing programs is the driving force behind fundamental change within schools; that the guarantee of rewards or the threat of sanctions is essential to promote high-quality teaching and student achievement. However, just as some have praised the high-stakes aspect of testing programs as the lynch-pin of successful educational reform, others suggest that the rewards and sanctions tied to test performance limit the scope of classroom instruction and learning.

Given the increasing reliance on state testing programs to determine high school completion and the large number of students affected by these policies, the need for more research on how the consequences of state-mandated testing programs affect instruction and learning is compelling. Consequently, the purpose of the National Board on Educational Testing and Public Policy (NBETPP) study that is the focus of this report was to collect information from those who witness the effect of state-mandated testing firsthand: classroom teachers. Teachers are charged with implementing testing programs and policies but often have little influence on their formulation. By gathering the opinions of teachers on high-stakes testing and its impact on teaching and learning, this study gives voice to those who generally are greatly affected by but only marginally involved in the processes that lead to statewide testing programs.

BACKGROUND

State education policymakers have a long history of instituting testing programs in response to concerns about the quality of education students receive. Tests have consistently been viewed as a lever to change classroom practices and produce overall improvement in general education. The current emphasis on high-stakes testing resulting from standards-based reform efforts is largely an extension of three decades of testing, with a new emphasis on higher standards and greater academic achievement. While rejecting notions of minimal competency and basic skills common to testing programs during the 1970's and 80's, standards-based reform efforts were designed to encourage schools, teachers and students to excel and meet tougher academic challenges as prescribed by state curricular standards or frameworks. In large part, state tests were designed to measure student achievement of these outcomes and serve as indicators of school quality.

To raise teachers' and students' performance levels, consequences serving as rewards and sanctions have therefore been attached to test results. These rewards and sanctions vary in severity. The logical extension of their use maintains that as consequences become greater, so does their capacity to motivate educational change (see Kelleghan, Madaus, & Raczek, 1996, for a review of the motivational aspects of tests). How the consequences attached to test results affect instruction and student achievement has been the focus of substantial research. Generally, this research has found positive and negative effects of state testing programs, particularly those with high stakes attached.

While the use of high-stakes testing is becoming more common, the landscape of state testing programs remains quite varied. The research conducted on the implementation and impact of state testing systems reflects this cross-state variability. Studies have been largely unsystematic and have involved testing programs with different stakes levels or testing formats (i.e. multiple-choice or performance-based). Research has also been inconsistent with regard to the grade level and content area at the focus of the study. But even though studies have varied in substantial methodological ways, they have generally been consistent with regard to the topics of interest. For example, most have focused on the effects of these tests on instruction with regard to what is taught, and how it is taught and assessed. Research efforts have also typically examined the role of test preparation and the relationship between the state test and the content standards, and some have addressed the psychological impact on the morale and motivation of teachers and students (see for example Firestone, Mayrowetz, & Fairman, 1998; Haney, 2000; Hoffman, Assaf, & Paris, 2001; Jones et al., 1999; Koretz, Mitchell, Barron & Keith, 1996; Koretz, Stecher, Klein, & McCaffrey, 1994; Lane, Parke, & Stone, 1998; McMillian, Myran, & Workman, 1999; Smith, Nobel, Heinecke et al., 1997; Stecher, Barron, Chun, & Ross, 2000; Stecher, Barron, Kaganoff, & Goodwin, 1998).

Impact on Instructional Content

Teachers make many decisions about what to teach, and how. One large area the vast majority of research has targeted is the influence of the state test on the focus of instruction and pedagogical methods. The results suggest that as stakes increase the curriculum will narrow to closely resemble the content sampled by the test (Corbett & Wilson, 1991; Madaus, 1998, 1991; Smith, 1991). More recent state-level studies report similar findings; that is, teachers are giving greater attention to tested content areas. For example, more than 80% of the 722 Virginia teachers surveyed indicated that the state Standards of Learning (SOL) test had affected their instruction (McMillan, Myran, & Workman 1999), leading the study authors to conclude that "teachers are placing greater emphasis on covering the content of the SOL" (p. 10).

Increased attention to tested content has often led to decreased emphasis on non-tested areas. A study in Arizona reported that teachers placed less emphasis on non-tested subjects such as social studies and science, while giving greater attention to the tested subject areas of English and math (Smith et al., 1991). In Kentucky, 87% percent of teachers surveyed agreed with the statement that the Kentucky Instructional Results Information Systems (KIRIS) "has caused some teachers to de-emphasize or neglect untested subject areas" (Koretz, Barron, Mitchell & Stecher, 1996, p. 41).

In the state of Washington, teachers' views corroborate this trend. Stecher et al. (2000) found that elementary teachers had increased instructional time spent on tested subjects and decreased time devoted to non-tested content in response to the Washington Assessment of Student Learning (WASL). The researchers found that the fourth grade teachers involved in the study spent 63% of their instructional time on tested areas (e.g. reading, writing and mathematics). Teachers in North Carolina also reported that non-tested curricular areas received minimal attention (Jones et al., 1999). Herman & Golan (n.d.) found that in addition to emphasizing tested content, teachers may alter the sequencing of their curriculum to ensure that they cover content most likely to appear on the state test.

Impact on Instructional Strategies

While research evidence strongly suggests that state tests often lead to increased emphasis on tested content areas, often at the expense of non-tested subjects, the impact of the test on the modes of instruction seems to depend on the format of the state test. Some research suggests that greater instructional emphasis is placed on higher-level thinking skills, particularly when state tests require written responses. For example, the majority of writing teachers surveyed in Kentucky indicated that the KIRIS writing portfolios had a positive effect on writing instruction (Stecher et al., 1998). Similarly, researchers involved in a previous study in Kentucky found that 80% of teachers reported increasing instructional emphasis on problem solving and writing as a result of the portfolio-based state test (Koretz et al., 1996a).

Stecher et al. (2000) found that instructional methods did not necessarily change in response to state testing; however, the frequency with which teachers used certain methods did change. For example, mathematics teachers reported increased use of open-ended questions and more often had students provide written explanations of the thought processes involved in their problem solving. Further, a majority of the writing teachers in the same study indicated that they had at least moderately changed their instruction methods. However, in Virginia, which administers a series of predominantly multiple-choice end-of-course exams, McMillian, Myran and Workman (1999) found that the state test had a greater impact on the content and pace of instruction than on the "mode of instruction." In addition, a study of Maryland and Maine that examined classroom practices led Firestone et al. (1998) to conclude that while teachers were aligning instruction with the state test, they were less likely to make changes in instructional methods.

Pressure on Teachers to Improve Student Performance

The pressure to respond to increased demands of the state test often requires teachers to place more emphasis on preparing students specifically for that test. In Maryland, 88% of teachers surveyed felt they were under "undue pressure" to improve student performance (Koretz et al., 1996b). An even larger proportion, 98%, of Kentucky teachers when asked the same question responded similarly (Koretz et al., 1996a). Increased emphasis on test preparation is one of the possible results of the pressure on teachers to improve student performance. Of the 470 elementary teachers surveyed in North Carolina, 80% indicated that "they spent more than 20% of their total instructional time practicing for the end-of-grade tests" (Jones et al., 1999, p. 201). Similarly, a survey of reading teachers in Texas revealed that on average teachers spent 8 to 10 hours per week preparing students for the Texas Assessment of Academic Skills (TAAS) (Hoffman, Assaf, & Paris, 2001). The most common test preparation activities reported by Texas teachers included demonstrating how to mark the answer sheet correctly, providing test-taking tips, teaching test-taking skills, teaching or reviewing topics that will be on the test, and using commercial test-preparation materials and tests from previous years for practice (Hoffman, Assaf, & Paris, 2001, p. 6).

One concern stemming from the reported emphasis on test preparation activities centers on the credibility or accuracy of test scores as a measure of student achievement. Specific test-preparation activities, coaching, and instruction geared towards the test can yield scores that do not agree with other, independent measures of the same content or skills (Haladyna, Nolen, & Haas, 1991; Koretz, Linn, Dunbar, & Shepard, 1991; Madaus, 1988; Smith, 1991). For example, 50% of Texas teachers surveyed did not think that the rise in TAAS scores "reflected increased learning and high-quality teaching" (Hoffman, Assaf, & Paris, 2001, p.8). Thus student performance on highly consequential tests may not be a credible or accurate measure of student achievement; specific test preparation may have corrupted the indicator, that is the state test results.

Impact on Motivation and Morale

Although intended to motivate teachers and students to reach higher performance levels, the high-stakes nature of state testing programs can have quite the opposite effect. With regard to teachers, researchers have cautioned that placing a premium on student test performance has led to instruction that is focused primarily on test preparation, thus limiting the range of educational experiences and reducing the instructional skills of teachers (McNeil, 2000; Smith, 1991). Studies also indicate that high-stakes assessments increase stress and decrease morale among teachers (Barksdale-Ladd & Thomas, 2000; Smith, 1991). According to Jones et al. (1999), more than 77% of the teachers surveyed indicated decreased morale; in addition, 76% reported that teaching was more stressful since the implementation of the North Carolina state-testing program. Similar results were found in Kentucky and Maryland. Over half of the Maryland teachers and about 75% of Kentucky teachers indicated that morale had declined as a result of the state test (Koretz et al., 1996a, 1996b). In addition, 85% of teachers surveyed by Hoffman, Assaf, and Paris (2001) agreed with the statement "Some of the best teachers are leaving the field because of the TAAS," suggesting that the emphasis on the TAAS was harmful to teaching.

While some research identified potentially harmful effects of high-stakes testing on the morale and professional efficacy of teachers, other studies identified similar concerns about students (Barkesdale-Ladd & Thomas, 2000). Increased anxiety, stress, and fatigue are often seen in these programs and can have detrimental effects on student performance. Of the teachers surveyed in North Carolina, 61% reported that their students were more anxious as a result of the state test (Jones et al., 1999). Similarly, one-third of teachers surveyed in Kentucky indicated that student morale had declined in response to the KIRIS (Koretz et al., 1996a).

Even though the rewards and sanctions attached to test results may spur many students to achieve and even excel, they may drive others out of school. If students do not believe that the opportunity for success exists, the motivating force of the rewards or sanctions will be small (Kellaghan, Madaus, & Raczek, 1996). Students who view passage of the test as an insurmountable barrier may give up and drop out of high school. In addition to research involving teachers' perceptions, empirical studies have shown that the use of high-stakes tests is associated with increased dropout rates (Haney, 2000; Reardon, 1996). This finding is especially disconcerting since initial passing rates on state exit exams are lower for minority students, students with disabilities, English as a Second Language learners and students from low socio-economic levels (Center on Education Policy, 2002).

Teachers' Views on Accountability

The results of state tests not only provide information about the progress of individual students; they are often aggregated to establish a measure to evaluate school and district performance. Schools face sanctions for poor student performance on state tests in at least 20 states (Quality Counts, 2002). They not only risk losing accreditation if students perform poorly, but also face funding losses and even the threat of a state takeover. Currently 18 states offer schools financial incentives for high or improved test scores (Quality Counts, 2002). Many policymakers believe that holding both schools and students accountable for test performance will produce fundamental positive educational change (Heubert & Hauser, 1999).

Most research studies on state testing programs have focused on the effects on classroom practices and have linked changes in instructional methods and content emphasis to the direct pressure to improve test scores. In addition, several studies have tapped into teachers' general perceptions of accountability. In North Carolina, 76% of the teachers surveyed "believed that the accountability program would not improve the quality of education in their state" (Jones et al., 1999, p. 202). Similarly, Barksdale-Ladd & Thomas (2000) discovered through interviews that teachers found their instruction "worse instead of better" as a result of the state test. In contrast, the majority of Kentucky and Washington teachers held positive views about the instructional impact of the state education reforms (Koretz et al. 1996a; Stecher et al., 2000). However, research conducted in Maine and Maryland suggested that teachers' perceptions of the stakes were not always consistent (Firestone, Mayrowetz & Fairman, 1998), suggesting that consequences attached to test performance can have a differential effect on schools within the same state. In other words, the intended effect of the rewards and sanctions tied to test performance may be mitigated by other factors (Firestone, Mayrowetz, & Fairman, 1998).

Overall, the research suggests that state tests have been a powerful influence on what gets taught in classrooms, and to a lesser extent on the methods of instruction. What does seem clear is that the evidence is mixed with regard to their success in improving the quality of education and their instructional value, given the added influence of high-stakes consequences. Research indicates both positive and negative results of state testing policies: greater emphasis on higher-level thinking skills and increased attention to writing is balanced by reported increases in stress, anxiety, and pressure to prepare for and perform on the state test. What has yet to be determined is whether the benefits of educational reform outweigh the unintended negative consequences, and how, if at all, stakes for students are influenced by stakes at the school level. The mixed and often contradictory results of state-level research highlight the need for a national look at the impact of state testing programs. That look is provided by this study.

METHODOLOGY

National Survey Development

An 80-item survey was used to elicit teachers' attitudes toward and opinions of state testing programs (see Appendix A). Many of the items in the survey were geared toward capturing the beliefs of teachers about the influence of their state's test on classroom instruction and student learning. The survey was based, in part, on other surveys used in Arizona (Smith, Nobel, Heinecke et al., 1997), Maryland (Koretz, Mitchell, Barron, & Keith, 1996), Michigan (Urdan & Paris, 1994) and Texas (Haney, 2000), as well as on the National Science Foundation (NSF) study of the Influence of Testing on Teaching Math and Science in Grades 4-12 (Madaus, West, Harmon, Lomax, & Viator, 1992) and a study of the Effects of Standardized Testing (Kellaghan, Madaus, & Airasian, 1980).

The survey consisted primarily of items in the form of statements or questions relating to standards-based education reform. A Likert response scale was used for most of these items to assess the intensity of opinion. Teachers were asked to indicate whether they "strongly agreed," "agreed," "disagreed," or "strongly disagreed." In addition to these closed-format items, the questionnaire also had an open-ended question that allowed teachers to write comments about the impact their state-mandated testing program had on their instructional practices and students' learning. The survey addressed the following topics:

- Information about state and district testing programs
- School climate
- Relationship of the mandated test to the state curriculum frameworks and standards
- Beliefs about teaching, learning, and assessment
- Classroom activities relating to instructional and testing practices
- Test preparation and administration
- Use and reporting of test results
- Professional development related to the state-mandated test
- Perceived effects of the state-mandated test

Former and current classroom teachers were involved in two field test administrations; their comments contributed to the refinement of the final survey items. The survey was administered during January-March 2001. The approach included a notification letter, a survey form, a reminder postcard, and an incentive to encourage participation in the study (Dillman, 2000). One follow-up mailing was conducted.

Sampling

In addition to answering the larger question of what opinions teachers hold of state-mandated testing programs, we were particularly interested in how teachers' attitudes differed depending on the consequences or stakes attached to test results. As each state is charged with its own educational policy development and implementation, state testing programs vary. Standards, tested content, item format, and consequences of test results differ from state to state. For example, Georgia, Massachusetts, Texas, and Virginia use test results to determine, in part, whether students are awarded high school diplomas and whether schools retain their accreditation (Quality Counts, 2002). Other states, such as Kentucky and Vermont, use student performance on the state test to hold schools, rather than students, accountable (Quality Counts, 2002). The first level of stratification used in our sampling design involved categorizing state testing programs according to the nature of the stakes attached to their test results.

The state classification process produced two general categories of stakes: (1) consequences for districts, schools, and/or teachers, and (2) consequences for students. Within each category, the severity of the stakes was classified as high, moderate, or low. The high-stakes category refers to state-regulated or legislated rewards and/or sanctions for schools, teachers, and/or students, such as whether or not (1) a student receives a high school diploma, (2) a student is promoted to the next grade, or (3) a school remains accredited (Heubert & Hauser, 1999). The low-stakes category included states with testing programs that had no known consequences attached to test scores. If the stakes for districts, schools and teachers and/or students were neither high nor low, states were placed in the moderate category. This included, for example, publicly disseminated test results in local newspapers, or including the results on students' school transcripts (Shore, Pedulla & Clarke, 2001). The classification of states was based on information found in state legislation, direct contact with state department of education personnel, and web sites at the time the survey was administered (January, 2001).

From this categorization, a nine-cell testing program matrix emerged (see Appendix B). However, based on the classification scheme, one cell remained empty and three cells contained only one state. Since it was cost-prohibitive to sample these three cells at the same rate as the other five, Iowa, Oregon and Idaho were excluded from the study. Once the states had been classified, 12,000 teachers were randomly selected to participate in the study. Teachers were also sampled according to the type of school in which they taught (elementary, middle and high), content area (e.g. English, math, science, social studies, and special education) and geographic setting of the school (i.e. urban and non-urban areas). Also incorporated in the sampling stratification was an oversample of Massachusetts teachers, of whom 1,000 were selected. This allowed the researchers to report specifically on that state (not part of this report). Table 1 presents the sampling frame, numbers and percentage of the teaching population within each stakes level, and numbers and percentage of teachers sampled to participate in the study.

All of the teachers in the sample were either regular classroom teachers who provided instruction related to core content areas (e.g., English, math, science and social studies) or teachers of special education students. The researchers assumed that teachers of core curriculum courses were most affected by state-mandated testing programs. Thus, educators who teach physical education, art and music or any other elective course were excluded from the sample. High school teachers were sampled at twice the rate of elementary and middle school teachers. Elementary school teachers included those who taught grades 2 through 5. The high school teachers were further categorized according to the subject they taught (i.e., English, math, science, social studies and special education). Within each of the cells, according to grade level and subject area, the sample is proportionally divided by location. This guaranteed that the proportion of teachers from both urban and non-urban (i.e. suburban and rural) areas was representative of the national population.

Table 1. Basic Sampling Frame

Consequences of State Testing Programs for Schools/Students	Total Number of Teachers	Percentage of Population	Percentage of Sample	Number of Teachers Sampled
High/High	1,488,226	56.83	18.33	2,200
High/Moderate	392,672	14.99	18.33	2,200
High/Low	238,417	9.10	18.33	2,200
Moderate/High	320,514	12.24	18.33	2,200
Moderate/Low	122,060	4.66	18.33	2,200
Massachusetts	57,097	2.18	8.33	1,000
Total	2,618,986	100.00	99.98	12,000

Source: Market Data Retrieval, 2000.

Description of Teacher Participants

Of the 12,000 teachers who received the national survey, 4,195 returned useable surveys, yielding a response rate of 35%.[1] Surveys were received from every state sampled (Iowa, Oregon and Idaho were excluded from the sample). The teachers varied widely with respect to personal characteristics and professional experience. The overwhelming majority were late-middle-aged females with considerable teaching experience. Approximately 67% of teachers who completed a survey were over 40 years old. Forty percent had more than 20 years of teaching experience. At the high school level, more English and math teachers responded than science, social studies or special education teachers. This was reasonable considering that most state testing programs focus on English and math as their primary tested areas (Quality Counts, 2002). Appendix D provides a detailed summary of the teachers in the sample and national comparison figures.

Data Analysis

Two sets of sampling weights were applied using the probability of selection from (1) the national teaching population and (2) the populations of the state testing programs to provide for a more accurate representation of the teaching force. The weights were the product of the inverses of the probability that the teacher would be selected from these populations and the response rate. The national population weights were applied when estimating teachers' responses nationwide, while state testing program weights were used when making comparisons among the different types of testing programs.

Descriptive statistics were calculated for each type of testing program and grade level, and frequencies were computed for the survey items. For the Likert items, and items with common response options, factor analyses were conducted to create scale scores and continuous variables that would permit significance-testing procedures, such as one-way and two-way analyses of variance. (Technical terms used through out the report are defined in Box 1). Significance testing was conducted at the individual item level using chi-square tests and tests of the difference in proportions as appropriate. Generally, a minimum percentage difference of 6 to 9% was needed for statistical significance at an alpha level of .001, the level used for all significance tests. In addition, results in some sections of the report were standardized and are graphically presented relative to a comparison group.

Generalizability of Findings

In comparison with the national population (see Appendix D), the teachers who completed the NBETPP survey were comparable in terms of age, race/ethnicity, the type of school in which they worked (elementary, middle or high school) and teaching experience. The similarity of the sample's demographics to that of the national population gives us confidence in our ability to generalize the results to the national teaching force. It is important to note the evolutionary nature of state testing programs. Since this national survey was administered, state testing programs have reached different points in their implementation. They may have changed substantially. Thus, specific state classifications made at the time of the study may not reflect the current situation. The findings about various stakes levels, however, should generalize to states that have those stakes levels now.

Organization of the Report

The results are organized by topic area, and within each topic are reported by stakes level of the testing program and by school level. To avoid verbosity, abbreviations are used for the five types of testing programs. As discussed previously, states were classified into five testing program categories along two dimensions: (1) stakes for districts, schools, and/or teachers, and (2) stakes for students. Testing programs that have high stakes for districts, schools, and/or teachers and high stakes for students are referred to as H/H; similarly, states with high stakes for districts, schools, and/or teachers and moderate stakes for students are referred to as H/M. The abbreviation H/L is used for states with high stakes for districts, schools, and/or teachers and low stakes for students; M/H is used for moderate stakes for districts, schools, and/or teachers and high stakes for students. Last, the abbreviation M/L is used in reference to testing programs that have moderate stakes for districts, schools, and teachers and low stakes for students. The main topic areas for the results section include:

- School climate
- Pressure on teachers
- Alignment of classroom practices with the state test
- Perceived value of the state test
- Impact on the content and mode of instruction
- Test preparation and administration
- Unintended consequences of the state test
- Use of test results

The final chapter summarizes the results, highlighting comparisons across the various types of testing programs, grade levels, and combinations of stakes and grade levels.

Box 1: Key Terms and Definitions

Scales

A scale is a subgroup of items in a questionnaire that measures one variable (or factor) in the survey. Survey questionnaires often use multiple items to measure a variable or factor. When designing a questionnaire, researchers need to decide the number of topics they want to cover, i.e. the number of variables on which they want to collect information. Multiple items are then written to measure each of these variables. A survey questionnaire usually consists of more than one scale.

Factor analysis

Empirically, factor analysis is a classical statistical procedure used to group items into scales. When designing questionnaires, researchers conceptually identify the factors to be covered (e.g. school climate, pressure on teachers) and write items to measure each factor. Results from factor analyses are expected to match the conceptual design. Theoretically, items measuring the same factor or variable should be highly correlated with each other, while the correlation between items measuring different factors should be much lower. Factor analysis capitalizes on this differential correlation pattern and groups items into different clusters (scales).

Cronbach's alpha

Cronbach's alpha, usually reported in survey studies, indicates the reliability of a scale (which consists of a number of items) in measuring a particular factor. Since all measurement involves error, all scores consist of two parts: the effect of the variable itself (the true score) and the effect of the error. Conceptually, Cronbach's alpha is the ratio of variation due to the true score to the total variation in the score. Theoretically, the value of Cronbach's alpha ranges between 0 and 1, with larger values indicating higher reliability, i.e. less measurement error. Restated, Cronbach's alpha indicates how homogeneous the items are in a scale that supposedly measures a single factor (internal consistency).

Standard deviation

A standard deviation is an index describing the amount of variation in a measure or variable. It takes into account the size of the sample, and the difference between each observation and the sample mean. Standard deviations are in the same unit of measurement as the original variable. Large standard deviations indicate greater heterogeneity in the sample, while small standard deviations indicate more homogeneity. In many natural phenomena, where the distribution approximates a normal distribution, about 68% of the cases lie within the range of one standard deviation below to one standard deviation above the mean, and roughly 95% of the cases lie within the range of two standard deviations below to two standard deviations above the mean.

One-way analysis of variance (One-way ANOVA)

In survey studies, participants may respond to items differently, and therefore there is variation in the responses, technically known as variance (variance is the square of the standard deviation). Often, the variation in responses may be related to who is responding, i.e. the group membership of the respondents. For example, teachers from different states may respond to a scale differently, indicating a between-group effect or difference in the response pattern. We would also expect there to be variations within each group simply because people are different even within the same state. In one-way ANOVA, we want to find out whether the between-group variation is significantly larger than the within-group variation, thus providing evidence of a group membership effect — i.e., respondents' group membership affecting their responses to a survey question.

Two-way analysis of variance (Two-way ANOVA)

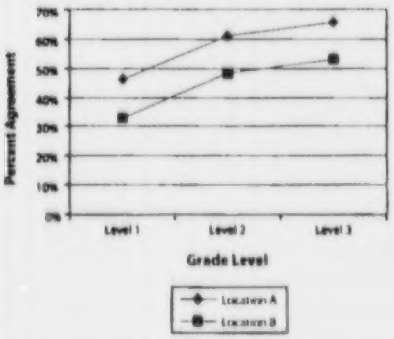
More often than not people are members of more than one group. For example, teachers can be identified by the location where they teach (e.g. the state), and also by the grade level they teach. It is possible that both location and grade level affect teachers' responses. Since two categories are involved, two-way analysis of variance is used to examine the effects of membership in each category (and in the combination of the two categories).

Main effect and interaction effect

Suppose it is found that both teaching location and grade level have an effect, known as main effects, on teachers' responses. Teachers in Location A are more positive on a measure than teachers in Location B. In Figure 1, the line representing Location A is above the line representing Location B. Further, teachers at lower grade levels are less positive on a measure than teachers at higher grade levels. In Figure 1, this pattern is clear: in both locations the score of grade level 2 is higher than that of grade level 1, and grade level 3 is higher than that at grade level 2.

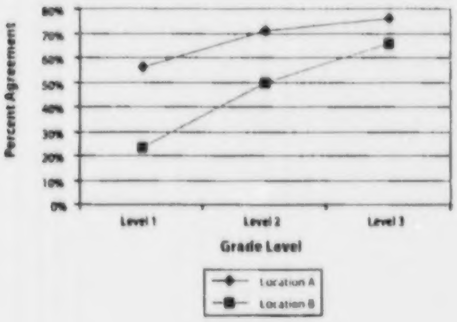
Box 1

Figure 1:
Main Effect in Two-way ANOVA



In addition to main effects, there may often be interactions between them (in this case location and grade level). We see in Figure 2 that the two lines representing the two locations are not parallel. Although the score in location A is still higher than in location B, the difference between the two is not consistent across grade levels: it is larger at lower grades smaller at higher grades. In this case, we say there is an interaction effect between location and grade level because both dimensions must be discussed in conjunction in order to explain the pattern of differences adequately.

Figure 2:
Main Effect and Interaction Effect in Two-way ANOVA



Chi-square and standardized residuals
Chi-square tests are used to find out whether observed frequencies on a certain variable across different groups differ significantly from expected frequencies. In a survey, respondents may agree or disagree with a statement; chi-square tests are used to find out whether different groups exhibit differences in their percentages of agreement to the statement.

Standardized residuals quantify the discrepancy between expected and observed values, indicating the direction (whether the observed value is larger or smaller than the expected value) and the magnitude of the discrepancy in standardized forms. If a chi-square test finds a significant overall difference across groups, standardized residuals help identify the cells where those differences occur.

Comparison of proportions
In survey studies, proportions (or percentages) of responses are often reported. As mentioned earlier, all measurement involves error; so too with proportions. When proportions are reported for two groups, direct comparisons can be easily made simply by calculating the difference between them. However, a test of statistical significance is needed to determine whether the calculated difference is due to chance, i.e. random fluctuation, or is large enough to be considered more than a chance difference.

Effect size
When more than one group is involved in a study, we often compare the groups. One common practice, for example, is to compare the means by looking at the difference. However, the interpretability of this difference depends on how variable the original scores are. A difference of 50 points between two group means must be interpreted very differently if the standard deviation for the measure is 1000 than if it is 100. By using effect size, the difference between two groups is represented as a proportion of the standard deviation of a reference group, and thus standardizes the difference. In the example above, the effect sizes would be .05 (50/1000) and .5 (50/100). In this way, we can see that the difference of 50 in the latter instance is much greater than the difference of 50 in the former instance. Effect sizes allow for direct comparison because they are on a common standardized scale. The interpretation of the magnitude of effect sizes depends on the situation. According to Cohen's criterion, an effect size of .25 is considered small, .5 medium, and 1.0 large. In practice, effect sizes of over half a standard deviation are rare (Mosteller, 1995). In the NBETPP survey study, graphs are used in many sections to illustrate the effect sizes of responses across different groups; these tend to range between .2 and .8 for most items.

RESULTS OF THE NATIONAL TEACHER SURVEY

I. Impact on School Climate

In order to determine whether teachers working in different types of testing environments had different perceptions of how the state test affected the atmosphere within schools, we examined several items together. Factor analysis results indicated that eight survey questions clustered together around teachers' perceptions of school climate to form a common scale (see Appendix E, Table E1). The school climate scale comprised the following items:

- My school has an atmosphere conducive to learning. (Item 36)
- Teachers have high expectations for the in-class academic performance of students in my school. (Item 34)
- The majority of my students try their best on the state-mandated test. (Item 32)
- Student morale is high in my school. (Item 26)
- Teachers have high expectations for the performance of all students on the state-mandated test. (Item 17)
- Many students are extremely anxious about taking the state-mandated test. (Item 33)
- Students are under intense pressure to perform well on the state-mandated test. (Item 41)
- Many students in my school cheat on the state-mandated test. (Item 51)

Overview of School Climate

The eight items of the school climate scale were coded so that higher values for individual items represented more positive perceptions of school climate. Items were initially coded 1 for "strongly disagree," 2 for "disagree", 3 for "agree" and 4 for "strongly agree." In order to create a balanced survey, items were positively and negatively worded; negatively worded items were then reverse-coded to maintain consistency in interpreting the scale score results.

In order to compare groups on the scale, scores were computed by averaging responses to the eight survey items. A two-way analysis of variance — stakes (H/H, H/M, H/L, M/H, M/L) by school level (elementary, middle, high) — was conducted to determine whether differences in means were statistically significant. The results of the statistical tests are presented in Appendix E, Table E2. The main effect for school type was significant; however, the main effect for stakes level and the effect of the interaction between stakes level and school type were not significant at alpha = .001. In other words, differences in teachers' opinions regarding school climate depended on the type of school rather than the type of state testing program. Consequently, teachers' views on their school's atmosphere were similar regardless of the consequences or stakes attached to the state test results.

Table 2 presents the mean for each school level on the school climate scale; higher means represent greater agreement or more positive perceptions of school climate. The mean values suggest that elementary teachers report the most positive school atmosphere (mean = 2.85) and high school teachers the least positive (mean = 2.73). However, the mean score for all of the stakes-level configurations places each group between "disagree" and "agree," thus suggesting that teachers generally maintained neutral views on the atmosphere in their schools.

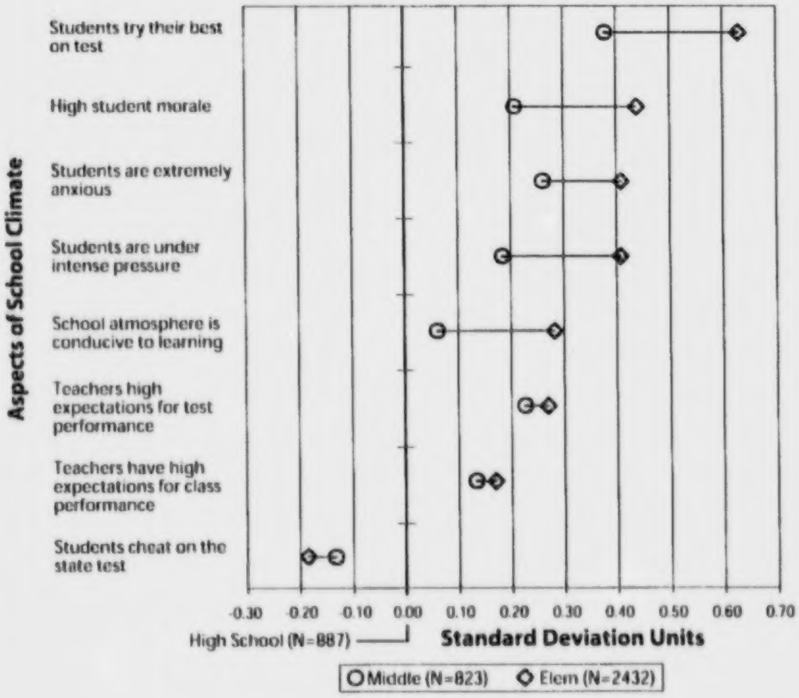
Table 2. Means on the School Climate Scale by School Type

School Type	N	Mean	SD
Elementary	2,476	2.85	.32
Middle	956	2.82	.32
High	735	2.73	.33

In order to provide a uniform method of comparison, mean scores on the eight items for each school type were standardized. As a result, the mean for high school teachers on the scale is 0 for each item, and the high school group serves as baseline or a point of comparison. The responses of elementary and middle school teachers are represented in standard deviation units relative to those of high school educators. Figure 3 shows how elementary and middle school teachers' responses deviated from those of teachers in high schools. The magnitude of positive and negative values indicates the degree of deviation. For example, smaller proportions of elementary and middle school teachers reported that students in their school cheat on the state-mandated test than did high school educators, as indicated by the negative values of the standard deviations (see Figure 3). In addition to the positive or negative value, the larger the magnitude of the deviation, the greater the departure or difference in the mean score for each item.

As Figure 3 illustrates, elementary and middle school teachers' responses are generally quite different from those of high school educators for some items, as indicated by the magnitude of standardized effect sizes (the largest range from .41 to .64). For the majority of items that make up the scale, the type of school has a substantial influence on teachers' responses. The responses of elementary school teachers' differed from those of high school practitioners on items that addressed effects on school atmosphere. For example, the proportion of elementary school teachers who reported that students in their school tried their best on the state test is almost two-thirds of a standard deviation greater than that of high school teachers. While Figure 3 illustrates substantial departures in perceptions related to school climate, it also shows similarities among teachers' responses regarding expectations for students' in-class performance, as shown by the small standardized deviation values or effect sizes (elementary = .17, middle = .13). As suggested in Figure 3, both positive and negative perceived effects of the state test on schools' atmosphere are more pronounced at the elementary school level.

Figure 3.
School Climate: Agreement of Elementary and Middle School Teachers vs. High School Teachers



Item-Level Results by Stakes Level

In order to explore further teachers' general views on school climate and their perceptions of how the state test influences their school's atmosphere, this section discusses individual survey items by stakes level. Even though there were no significant differences at the scale level, there were some at the item level. Teachers' responses across the five testing program configurations were similar regarding overall characteristics of school climate. For example, about 9 in 10 teachers within each type of testing program indicated that their school has an atmosphere conducive to learning (see Table 3). In addition, roughly similar proportions of teachers across stakes levels reported that teachers in their school have high expectations for the in-class performance of students. Teachers were also in agreement about their expectations of students' performance on the state test, even though these were generally lower than those for in-class achievement. Approximately 65% of teachers within each type of testing program indicated they held high expectations for students' performance on the state test. Last, very few — roughly 5% of teachers across the stakes levels — said that students in their school cheat on the state test.

Table 3. Views on School Climate: Percent Agreement by Stakes Level[1,2]

School Climate Related Items	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
My school has an atmosphere conducive to learning.	92	91	93	92	92
Teachers have high expectations for the in-class academic performance of students in my school.	91	92	93	91	89
The majority of students try their best on the state-mandated test.	84	77	78	85	77
Student morale is high in my school.	65	72	72	72	79
Teachers have high expectations for the performance of all students on the state-mandated test.	67	64	65	63	63
Many students are extremely anxious about taking the state-mandated test.	80	76	70	83	72
Students are under intense pressure to perform well on the state-mandated test.	80	68	68	75	49
Many students in my school cheat on the state-mandated test.	3	5	4	3	6
Many students in my class feel, that, no matter how hard they try, they will still do poorly on the state-mandated test.	52	55	56	61	54

1. Shaded values indicate statistically significant percentage differences from the moderate/low category (alpha = .001).
2. The strongly agree and agree response categories were collapsed into general-agreement responses.

Teachers' responses diverged with respect to the school climate experienced by students. For example, significantly fewer teachers in H/H (65%), H/M, H/L and M/H (72% each) stakes states reported that student morale was high in their school than did those in M/L states (79%). Asked specifically about students and the state test, teachers from states with high stakes for students responded similarly. For example, far more teachers from H/H (84%) and M/H states (85%) than from M/L states (77%) reported that most of their students try their best on the state test. In addition, more teachers in H/H (80%) and M/H states (83%) than M/L teachers (72%) agreed that "students were extremely anxious about taking the state test." This opinion was especially intense for H/H (35%) and M/H teachers (37%), of whom over a third strongly agreed that students were extremely anxious.

Asked about general pressure on students rather than specifically test-related anxiety, teachers responded somewhat differently. While test-related anxiety seemed to result from the stakes for students, intense pressure to perform well on the state test seemed to be influenced by the stakes for both schools and students. For example, significantly greater percentages of teachers in H/H (80%), H/M (68%), H/L (68%), and M/H stakes (75%) states than in M/L states (49%) agreed that students were under intense pressure to perform well on the test. This opinion particularly resonated with teachers in H/H and M/H stakes, roughly a third — H/H (32%) and M/H (30%) — strongly agreed that students were under intense pressure.

Even though responses about the anxiety and pressure students varied, teachers' opinions about test-related motivation were consistent. Most teachers across stakes levels indicated that many students feel that, no matter how hard they try, they will still do poorly on the test; of H/H teachers 17% versus 9% of M/L teachers strongly agreed. The last several items examined together suggest that teachers perceived students to feel test-related pressure and anxiety to some degree, and that most of them, regardless of the type of testing program, recognized that many students doubt they can succeed on the state test.

Item-Level Results by School Type

Item-level results according to school type (elementary, middle, high) show that teachers' perceptions of school climate vary, with substantial differences between elementary and high school teachers' opinions (see Figure 3). Generally elementary teachers hold more positive opinions about their school's atmosphere than do high school practitioners. While an overwhelming proportion of all teachers reported that their school's atmosphere is conducive to learning, more elementary teachers (95%) held this view than did middle (87%) or high school educators (87%). Table 4 presents the item-level results by school type. In addition, more elementary than high school educators indicated that teachers in their school held high expectations for student performance. For example, 69% of elementary educators compared with 55% of high school educators maintained that teachers have high expectations for student performance on the state test. Slightly larger percentages of elementary (92%) than high school teachers (88%) reported that teachers at their school have high expectations for in-class performance.

Teachers' response patterns for student-focused items also varied according to school type. The proportion of teachers reporting that student morale was high in their school was significantly different across school levels. More elementary (73%) than middle (65%) or high school teachers (56%) so reported. Items targeting psychological or behavioral effects of the state test showed similar disparities. Elementary teachers reported in significantly larger numbers (89%) than middle (79%) or high school teachers (66%) that most students try their best on the state test. Elementary teachers were also more likely to indicate that students felt pressure and anxiety as a result of the state test. Eighty-two percent of elementary teachers agreed or strongly agreed with the statement, "Many students are extremely anxious about taking the state-mandated test," while 77% of middle and 69% of high school educators held that view. A similar pattern emerged with respect to test-related pressure that students feel. Seventy-nine percent of elementary teachers indicated that students feel intense pressure to perform well on the state-mandated test, as compared with 73% of middle and 66% of high school teachers.

Elementary teachers were less likely to report incidents of cheating. Even though the incidence was low overall, significantly more high school (7%) than middle (3%) or elementary teachers (3%) reported that many students in their schools cheat on the state test. Similarly, elementary teachers were less likely to agree or strongly agree with the statement, "Many students in my class feel that, no matter how hard they try, they will still do poorly on the state-mandated test." Greater percentages of high school (62%) than elementary school teachers (49%) reported that students feel their efforts to succeed on the state test to be ineffective. Even though elementary teachers perceived students to be more anxious and under greater pressure than did middle or high school teachers, they were more likely to report that students tried their best and believed they could be successful on the state test.

Table 4. Views on School Climate: Percent Agreement by School Type[1,2,3]

School Climate Related Items[1]	School Type		
	Elementary	Middle	High
My school has an atmosphere conducive to learning.	***95***	87	87
Teachers have high expectations for the in-class academic performance of students in my school.	92	91	88
The majority of students try their best on the state-mandated test.	***89***	79	66
Student morale is high in my school.	***73***	65	56
Teachers have high expectations for the performance of all students on the state-mandated test.	69	67	55
Many students are extremely anxious about taking the state-mandated test.	82	77	69
Students are under intense pressure to perform well on the state-mandated test.	***79***	73	66
Many students in my school cheat on the state-mandated test.	3	3	7
Many students in my class feel, that, no matter how hard they try, they will still do poorly on the state-mandated test.	***49***	59	62

1. Shaded values indicate statistically significant percentage differences from the high school category at alpha = .001.

2. Italicized values indicate statistically significant percentage differences between the elementary and middle school results at alpha = .001.

3. The strongly agree and agree response categories were collapsed into general-agreement responses.

Summary

The results suggest that teachers' opinions of school climate depend largely on the type of school in which they work. The data show that as grade level increased, perceptions of school climate became more negative. More elementary than high school teachers maintained high expectations for students' in-class achievement (92% v. 88%). Both middle (67%) and elementary teachers (69%) reported significantly more often than high school teachers (55%) that they held high expectations for student performance on the state test. At the same time, significantly more elementary than high school teachers indicated that students were anxious and under intense pressure as a result of the state test. In other words, the psychological impact was perceived to be greater at the elementary level, yet this did not seem to negatively affect the general atmosphere of the school. Conversely, high school educators reported the general climate of the school to be less positive than those in elementary schools, yet they also reported lower levels of test-related pressure and anxiety in students. These results seem counterintuitive, particularly since the most severe sanctions for poor test performance usually occur at higher grade levels where test scores may be used to make decisions about grade promotion or high school graduation.

BEST COPY AVAILABLE

Results varied less across stakes levels. Only when items focused specifically on students did teachers' responses significantly differ. Teachers from high-stakes states were more likely to report that students were under intense pressure to perform well on the state test than were M/L teachers. In addition, many more H/H and M/H teachers indicated that students were extremely anxious about taking the state test than did teachers in M/L states. For expectations of students' in-class achievement and performance on the state test, teachers' responses across stakes levels were similar. Generally, teachers' perceptions of students' test-related anxiety and pressure seemed not to affect their expectations of student performance or perceptions of school climate. In other words, even though teachers reported students to be under pressure and anxious about the test, they maintained high expectations, particularly of students' in-class achievement, and remained positive about the general atmosphere within their school.

II. Pressure on Teachers

Within the context of school climate, teachers also feel pressure as result of the state test. A primary purpose of state testing programs with high stakes attached is to motivate administrators, teachers, and students to meet established curricular standards and increase academic achievement. Given the varied nature of accountability systems nationwide, it is unclear what combination of stakes for districts, schools, teachers, and students maximizes the benefits of standards-based reform without exerting undue pressure to prepare students for the state test. In an effort to gain insight into this issue, we asked teachers a series of questions related to pressure and how feelings of test-related pressure affect classroom instruction and their profession.

Overview of Pressure on Teachers

To explore how teachers working in different testing environments experience test-related pressure, we examined several items together. Factor analysis results indicated that eight survey questions clustered together around test-related pressure to form a common scale (see Appendix E, Table E3). That scale comprised the following items:

- Teachers feel pressure from the district superintendent to raise scores on the state-mandated test. (Item 21)
- Teachers feel pressure from the building principal to raise scores on the state-mandated test. (Item 47)
- Teachers feel pressure from parents to raise scores on the state-mandated test. (Item 37)
- Administrators in my school believe students' state-mandated test scores reflect the quality of teachers' instruction. (Item 49)
- The state-mandated testing programs lead some teachers in my school to teach in ways that contradict their own ideas of good educational practice. (Item 44)

- There is so much pressure for high scores on the state-mandated test teachers have little time to teach anything not on the test. (Item 39)
- Teacher morale is high in my school. (Item 13)[2]
- Teachers in my school want to transfer out of grades where the state-mandated test is administered. (Item 43)

Items were initially coded 1 for "strongly disagree," 2 for "disagree", 3 for "agree" and 4 for "strongly agree." Negatively worded items were then reverse-coded to maintain a common interpretation of the scale. Scale scores were computed by averaging responses to the eight items. Higher means on the scale indicate stronger feelings of pressure associated with the state test.

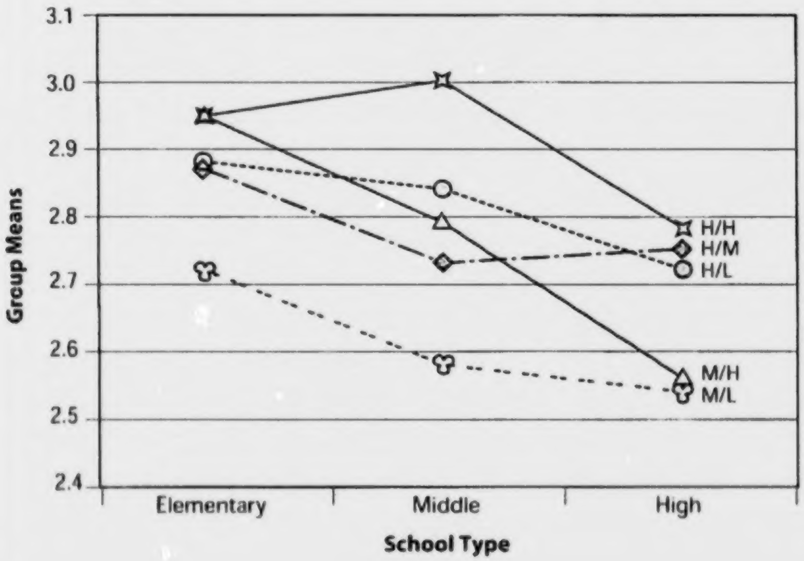
A two-way analysis of variance, comparing stakes (H/H, H/M, H/L, M/H, M/L) and school types (elementary, middle, high) was conducted for the pressure scale to determine whether mean differences on the scale were statistically significant. The results are presented in Appendix E, Table E4. Both of the main effects and the interaction effect were significant at alpha = .001. In other words, the test-related pressure teachers experience is linked to the combination of the type of school in which they work and the consequences associated with their state's testing program.

Table 5 presents the mean for each stakes level and school type on the pressure scale, while Figure 4 graphically displays these results. As shown in Figure 4, test-related pressure varies across grade levels within the same type of testing program, suggesting that stakes attached to test results have a different impact at the elementary, middle and high school level. For example, in both the H/H and H/L categories elementary and middle school teachers have similar mean scores on the pressure scale; however, within the other three stakes groups elementary teachers have larger means than middle or high school teachers, often substantially larger. Similarly, in the H/M and M/L categories, middle and high school teachers reported experiencing comparable amounts of test-related pressure, while in the remaining three stakes-level categories (H/H, H/L, H/M), middle school practitioners indicated feeling greater pressure than did high school teachers.

Table 5. Means on the Pressure Scale by Stakes Level and School Type

School Type	Stakes Level					
	H/H	H/M	H/L	M/H	M/L	Overall
Elementary	2.95	2.87	2.88	2.95	2.72	2.88
Middle	3.00	2.73	2.84	2.79	2.58	2.79
High	2.78	2.75	2.72	2.56	2.54	2.68
Overall	2.93	2.81	2.85	2.86	2.66	

Figure 4.
Pressure Scale Means: School Type by Stakes Level



We would not expect to see patterns of greater pressure on teachers at the lower grades, since the most severe consequences associated with state tests usually occur at the high school level. In these types of testing programs the stakes are much greater for high school students who must pass the test for graduation. While elementary or middle school students may be denied promotion to the next grade as a result of test performance, this sanction is less often imposed than those connected to high school graduation.

The pressure teachers experience as a result of the state test is influenced by the stakes attached to the test in combination with the grade taught. The results presented in Table 5 and Figure 4 clearly illustrate that with one exception, elementary teachers report significantly greater feelings of test-related pressure than teachers in the upper grades, particularly in states where stakes are highly consequential for schools and students.

Item-Level Results by Stakes Level

In order to explore further the relationship between the perceived influence of the state test and the pressure teachers feel to raise test scores and prepare students, this section discusses individual survey items related to pressure. Significantly more teachers in high-stakes states than in M/L states are reporting that they feel pressure from their district superintendent (92% vs. 84%) and their building principal (85% vs. 68%) to raise test scores. The similarity in the percentage of teachers so reporting in high-stakes states shows that this pressure is felt in either situation — high stakes for schools or high stakes for students (see Table 6). In contrast, teachers report feeling less pressured by parents. About 50% of teachers across stakes levels "felt pressure from parents to raise scores on the state test," suggesting that pressure from parents does not increase or decrease with the stakes attached to test results.

Table 6. Pressure on Teachers: Percent Agreement by Stakes Level[1,2]

Pressure-Related Items	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Teachers feel pressure from the district superintendent to raise scores on the test.	92	92	91	91	84
Teachers feel pressure from the building principal to raise scores on the test.	85	80	82	81	68
Teachers feel pressure from parents to raise scores on the state test.	55	52	52	58	54
Administrators in my school believe students' state-mandated test scores reflect the quality of teachers' instruction.	63	64	61	54	53
The state-mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good educational practice.	76	71	72	76	63
There is so much pressure for high scores on the state-mandated test teachers have little time to teach anything not on the test.	80	67	73	69	56
Teacher morale is high in my school.	43	51	53	47	46
Teachers in my school want to transfer out of the grade where the state-mandated test is administered.	38	29	40	39	18

1. Shaded values indicate statistically significant percentage differences from the moderate/low category (alpha = .001).
2. The strongly agree and agree response categories were collapsed into general-agreement responses.

Pressure to raise test scores plays out in the classroom. Seven in ten teachers in the four high-stakes categories reported that their state-mandated testing program has required them to deliver instruction that runs counter to their own ideas of good practice. More teachers in states with high stakes for students (76%) vs. M/L states (63%) agreed with this notion, suggesting that high stakes for students may contribute to a decline in what teachers view as pedagogically sound instruction. However, other factors may also be at work. The large proportion of teachers in the M/L category who reported that their state test leads them to teach in ways departing from good practice is noteworthy. According to these teachers, state policies with minimal sanctions for districts, schools, teachers, and students have negatively affected their instruction. Thus influences related to the implementation of state testing programs, regardless of the consequences, may affect what and how teachers teach.

Teachers in high-stakes states report far more often than M/L teachers that the pressure for high scores all but precludes their teaching material that does not appear on the state test. In H/H states, 80% of them reported feeling pressured to "teach to the test," in contrast to teachers in H/M (67%), H/L (73%), and M/H (69%) stakes states. According to teachers, the compounded impact of high stakes for districts, schools, teachers as well as students constrains their instruction.

While teachers from H/H states are the most likely to indicate feeling pressured to teach tested content, they are not more likely than those from H/L and M/H states to report that teachers in their school have transferred out of tested grades. The percentage of teachers from H/H (38%), H/L (40%), M/H (39%), and H/M (29%) stakes states who indicated that teachers at their school want to transfer out of those grades was considerably higher than for teachers in M/L states (18%). These results suggest that testing programs involving high stakes for either districts, schools, and teachers or for students, or both, contribute to teachers' desire to transfer into non-tested grades, especially since there was little disparity in teacher morale across stakes levels. This is particularly notable since a change in teaching position often requires a substantial time investment to plan for instruction that may involve different subject matter or targets different cognitive skills.

Item-Level Results by School Type

While teachers' opinions diverged noticeably between high-stakes and M/L stakes states, still greater variation is seen when the same items are examined across grade levels. As noted earlier, elementary teachers reported significantly greater feelings of test-related pressure than their middle and high school counterparts. This trend remains prominent, especially with regard to items that address the impact of test-related pressure on classroom instruction and professional status. Table 7 presents the test-related pressure items for each school type.

As was the case across stakes levels, greater pressure to raise test scores was felt from the district superintendent than from building principals or parents according to grade level. More elementary (84%) and middle school (85%) teachers felt this pressure from their principal than did high school teachers (76%). However, as with the stakes levels, there was no substantial difference across grade levels with regard to test-related parental pressure, suggesting that the parental pressure teachers experience depends less on the testing program or grade level than do the other pressure-related items. Still, half of the teachers in elementary, middle and high schools report experiencing some degree of parental pressure.

Table 7. Pressure on Teachers: Percent Agreement by School Type[1,2]

Pressure-Related Items	School Type		
	Elementary	Middle	High
Teachers feel pressure from the district superintendent to raise scores on the test.	93	92	85
Teachers feel pressure from the building principal to raise scores on the test.	84	85	76
Teachers feel pressure from parents to raise scores on the state test.	56	56	51
Administrators in my school believe students' state-mandated test scores reflect the quality of teachers' instruction.	63	63	56
The state-mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good educational practice.	78	73	67
There is so much pressure for high scores on the state-mandated test teachers have little time to teach anything not on the test.	79	77	61
Teacher morale is high in my school.	47	44	43
Teachers in my school want to transfer out of the grade where the state-mandated test is administered.	***43***	29	24

1. Shaded values indicate statistically significant percentage differences from the high school category at alpha = .001.
2. Italicized values indicate statistically significant percentage differences between the elementary and middle school results at alpha = .001.
3. The strongly agree and agree response categories were collapsed into general-agreement responses.

A substantial majority of teachers at each grade level indicated that state testing programs have led them to teach in ways that conflict with their ideas of sound instruction. This opinion was particular notable at the elementary level. Seventy-eight percent of elementary teachers as compared with 73% of middle and 67% of high school teachers held this view. In addition, many more elementary (79%) and middle (77%) than high school teachers (61%) indicated that there was "so much pressure for high scores on the state test that they had little time to teach content that did not appear on the test." Elementary teachers were almost twice as likely as high school teachers to suggest that teachers at their school wanted to transfer out of the grades in which the test was administered. These results may be partly a result of teachers being assigned to a grade at the elementary level while high school teachers may teach multiple grades within subject area departments. Although these results indicate that teachers at all grade levels feel significant pressure associated with the state test, elementary teachers are especially affected by heightened expectations to improve student performance. This may be because, unlike their counterparts, they have two or more tested areas to contend with.

Summary

The majority of teachers report substantial feelings of pressure related to the state-mandated test regardless of stakes level. However, the most acute pressure was felt by elementary school educators, particularly those working in states that have high stakes for students. Especially troubling is the widespread opinion that the pressure to raise test scores requires modes of instruction that are contrary to teachers' notions of good educational practice. High-stakes consequences for districts, schools, teachers, and students seem to intensify this view. Roughly 7 in 10 teachers in high-stakes states reported that their state test has negatively affected their instructional practice. Attaching high stakes to test results may, in their view, limit the quality of instruction. In addition, the highly consequential nature of state-testing programs may adversely affect the teaching profession by giving rise to a desire to transfer out of tested grades.

III. Alignment of Classroom Practices with the State Test

State-mandated testing programs have considerable influence on what happens in classrooms. State curricular frameworks and testing requirements affect teachers' daily decisions about content, lessons, and assessment of student learning. What is not clear, however, is how the consequences attached to test results shape the relationship between classroom practices and the state test. The results presented in this section focus on this relationship and how teachers' perceptions vary according to the stakes attached to the state test and the grade level they teach.

Overview of Alignment

To obtain an overview of how teachers working in different testing environments, view the impact of the state test on classroom activities, responses to survey items relating to alignment issues were examined collectively. Factor analysis results indicated that several survey questions clustered together around alignment issues to form a common scale (see Appendix E, Table E5). The alignment scale comprised the following items:

- My district's curriculum is aligned with the state-mandated test. (Item 9)
- The state-mandated test is compatible with my daily instruction. (Item 7)
- The state-mandated test is based on a curriculum framework that all teachers in my state should follow. (Item 10)
- My tests have the same content as the state-mandated test. (Item 50)
- The instructional texts and material that the district requires me to use are compatible with the state-mandated test. (Item 14)
- My tests are in the same format as the state-mandated test. (Item 42)

Items were coded 1 for "strongly disagree," 2 for "disagree," 3 for "agree" and 4 for "strongly agree"; consequently higher means represent a stronger association between classroom instruction and the state test.[3] Table 8 presents the mean for each stakes level on the alignment scale. Teachers in H/H (mean = 2.66) and H/L (mean = 2.71) stakes states have significantly higher mean values on the alignment scale than the other stakes levels. These results suggest that their classroom activities are more closely associated with the content and format of the state test than those of their counterparts working in other settings. However, the mean score for all of the stakes-level configurations places each group between "disagree" and "agree" on this scale, suggesting generally neutral opinions.

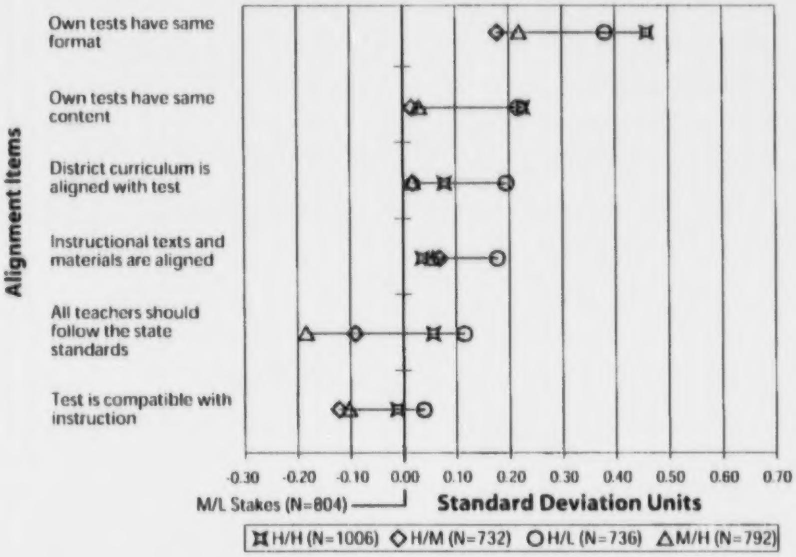
Table 8. Means on the Alignment Scale by Stakes Level

Stakes Level	N	Mean	SD
H/H	941	2.66	.50
H/M	783	2.56	.48
H/L	695	2.71	.43
M/H	839	2.54	.42
M/L	815	2.54	.43

In order to compare groups on the alignment scale, scores were computed by averaging responses to the six items. A two-way analysis of variance — stakes (H/H, H/M, H/L, M/H, M/L) by school type (elementary, middle, high) — was conducted to determine whether mean differences on the scale were due to something other than chance. The results are presented in Appendix E, Table E6. The main effect for stakes level was significant; however, the main effect for school type and the interaction effect of stakes level with school type were not significant at alpha = .001. In other words, teachers' views about alignment issues differed significantly by stakes level. However, there was no significant difference in scale means across the three grade levels.

Figure 5 presents an overall view of the six items that compose the alignment scale. The graph depicts teachers' responses by stakes level (H/H, H/M, H/L, M/L) versus the responses for the M/L group. The proportion of teachers agreeing with each item was transformed into standard deviation units relative to the responses of M/L teachers, thus allowing for more meaningful comparisons on a common scale. Figure 5 shows that the greatest departure from the M/L responses is related to the format of the state test. Teachers from H/H states are over .4 standard deviation units away from the 0 baseline, meaning that these teachers are far more likely to construct their own tests in the same format as that of the state test than those in M/L states. In addition, the negative standard deviation unit values for two items suggest that the instructional texts and materials for M/H and H/M teachers are less aligned with the state test. These teachers are also less likely to report that their state test is based on a curriculum that all teachers should follow than are M/L teachers.

Figure 5.
Alignment with the State Test: Agreement by H/H, H/M, H/L, and M/H, vs. M/L Stakes States



Item-Level Results by Stakes Level

In order to explore further the relationship between instruction and the stakes attached to state tests, this section discusses individual survey items related to the alignment of classroom practices with the state test. Standards-based reform was founded on the premise that if teachers teach the content standards, students will be prepared for the state test. Most teachers surveyed agreed with the statement, "If I teach to the state standards or frameworks, students will do well on the state-mandated test." However, significantly fewer teachers in H/H (54%), H/L (53%), and M/H (51%) states held this view than teachers in states with minimal consequences at the school or student level (M/L, 63%). In addition, significantly fewer teachers in M/H stakes states (48%) indicated that the state test was based on a curriculum that all teachers should follow. Table 9 presents a summary of results for the survey items related to standards and alignment with the state test.

Even though teachers from the H/H, H/L, and M/H testing programs were uncertain about how teaching to their state standards affected student performance, many of them indicated that they aligned their classroom tests with the content and format of the state test. Teachers in H/H (59%) and H/L (59%) states more often reported that their classroom tests have the same content as the state test than did their counterparts in states with low stakes for students (48%). Similarly, significantly more teachers in H/H (51%), H/L (47%) and M/H (40%) states indicated that they designed their own classroom tests to mirror the format of the state test than did teachers in M/L states (29%).

Table 9. Alignment with the State Test: Percent Agreement by Stakes Level[1,2]

Alignment-Related Items	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
If I teach to the state standards or frameworks, students will do well on the state-mandated test.	54	59	53	51	63
The state-mandated test is compatible with my daily instruction.	65	60	68	61	66
The state-mandated test is based on a curriculum framework that all teachers should follow.	60	53	63	48	57
My district's curriculum is aligned with the state-mandated testing program.	80	77	84	77	76
The instructional texts and materials that the district requires me to use are compatible with the state-mandated test.	59	60	65	60	57
My tests have the same content as the state test.	59	49	59	49	48
My tests are in the same format as state test.	51	38	47	40	29

1. Shaded values indicate statistically significant percent differences from the moderate/low category (alpha = .001).
2. The strongly agree and agree response categories were collapsed into general-agreement responses.

However, in states where the results of the test are highly consequential and apply only to districts, schools, and/or teachers, the responses suggest greater emphasis at the district level on supporting curricular alignment with the state test. For example, more teachers in H/L (84%) than in M/L stakes states (76%) indicated that their district's curriculum is aligned with the state test. These results do not suggest that curriculum is less aligned elsewhere; large majorities —roughly 75% of teachers across stakes levels — indicated that their district's curriculum was aligned with the state test. In addition, more teachers in H/L stakes states (65%) than M/L teachers (57%) reported that the instructional texts and materials required by the district were aligned with the state test. Regardless of the extent to which teachers are aligning the content and format of their classroom tests with those of the state test, they agree on the compatibility of the state test with their daily instruction. Roughly 60% of teachers across stakes levels agreed with the statement that "the state test is compatible with my daily instruction."

Item-Level Results by School Type

Individual item results highlight some differences in opinion across grade levels, even though there were no significant differences in the overall mean scale scores. Generally, larger percentages of elementary teachers indicated that their state test was based on a curriculum that all teachers should follow, and were more likely than their high school counterparts to report that they aligned their classroom assessments with the content and format of the state test (see Table 10). More elementary (60%) than high school teachers (53%) indicated that

BEST COPY AVAILABLE

their state test is based on a curriculum that all teachers should follow. In contrast, teachers' responses were similar across grade levels with regard to the impact of teaching to the standards or frameworks on students' test performance. Roughly 55% of teachers at each grade level indicated that if they aligned their curriculum with the state standards, students would do well on the state test. Slightly more teachers, approximately 60% at each grade level, reported that the state test was compatible with their daily instruction. However, significantly more elementary (58%) than high school teachers (50%) reported that the content of their classroom tests mirrored that of the state test. In addition, 49% of elementary and 48% of middle school teachers indicated that they constructed their tests in the format of the state test, while 38% of high school teachers so responded. Further, 58% of elementary and 59% of middle school teachers compared with 50% of high school teachers reported that their classroom tests had the same content as the state test. In contrast, a smaller proportion of elementary (56%) than high school teachers (66%) reported that the district's instructional texts and materials were compatible with the state test. These results suggest that instructional support in aligning the curriculum with the state test may be greater for high schools than elementary schools. Or perhaps, because of the structure of high schools, texts and materials were already compatible with the content of the state test at the onset of implementation.

Table 10. Alignment with the State Test: Percent Agreement by School Type[1,2,3]

Alignment-Related Items	School Type		
	Elementary	Middle	High
If I teach to the state standards or frameworks, students will do well on the state-mandated test.	54	55	56
The state-mandated test is compatible with my daily instruction.	64	66	62
The state-mandated test is based on a curriculum framework that all teachers should follow.	60	59	53
My district's curriculum is aligned with the state-mandated testing program.	78	83	81
The instructional texts and materials that the district requires me to use are compatible with the state-mandated test.	**56**	63	66
My tests have the same content as the state test.	58	59	50
My tests are in the same format as state test.	49	48	38

1. *Shaded and values indicate statistically significant percent differences from the high school category at alpha = .001.*
2. *Italicized values indicate statistically significant percent differences between the elementary and middle school results at alpha = .001.*
3. *The strongly agree and agree response categories were collapsed into general-agreement responses.*

Summary

Regardless of the consequences attached to the state test, a majority of teachers reported that their state test is based on a curriculum that all teachers should follow. About 60% of teachers in each type of testing program reported that their state test is compatible with their daily instruction. The data also show that a majority of teachers across stakes levels reported that the district's curriculum and required instructional texts and materials are aligned or compatible with the state test, with teachers in H/L testing programs being more likely than their peers in other states to report this alignment. Similarly, teachers indicated that classroom assessment practices closely resembled both the content and format of the state test; almost 60% of teachers in H/H and in H/L stakes states indicated that their tests had the same content. However, the influence of the stakes attached to the test was more noticeable in the format of classroom assessments. Roughly 50% of teachers in H/H and H/L and 40% of M/H teachers indicated that their tests were in the same format as the state test (see Table 9). With regard to school type, elementary teachers reported in significantly greater percentages than high school teachers that they aligned the content of their instruction and tailored classroom assessments to the state test. Generally, the impact of the state test on classroom assessments is more pervasive at the elementary than high school level. High school teachers reported more often that their curriculum, instructional texts, and materials were aligned with the state test so that the need to change classroom practices may not have been as great.

IV. Teachers' Perceptions of the State Test's Value

Standards-based reform efforts were designed to raise academic achievement. In an effort to measure student attainment of that goal, various forms of state tests were introduced. The value of the state test lies in its intended function to measure student achievement and serve as an indicator of school quality. In order to gain an understanding of what combination of stakes for districts, schools, teachers, and students makes the test valuable to teachers we explored its benefits and its capacity to fulfill its intended function — measure student achievement and school quality.

Overview of the Perceived Value of the State Test

In order to determine whether teachers working in different testing environments valued their state test differently, we examined several items together. Factor analysis results indicated that 13 survey questions clustered together around teachers' perceptions of the state test's value to form a common scale (see Appendix E, Table E7). These items focused on general perceptions of the value of the state test, the accuracy of the state test as a measure of student achievement, and media coverage of educational issues related to the state test. The perceived value scale comprised the following items:

- Overall, the benefits of the state-mandated testing program are worth the investment of time and money. (Item 11)

- Media coverage of the state-mandated test accurately reflects the quality of education in my state. (Item 23)
- Scores on the state-mandated test accurately reflect the quality of education students have received. (Item 15)
- The state-mandated test has brought much-needed attention to education issues in my district. (Item 40)
- The state-mandated test is as accurate a measure of student achievement as a teachers' judgment. (Item 8)
- The state-mandated test motivated previously unmotivated students to learn. (Item 20)
- The state-mandated test measures high standards of achievement. (Item 29)
- The state-mandated testing program is just another fad. (Item 16)
- Media coverage of state-mandated testing issues has been unfair to teachers. (Item 30)
- Media coverage of state-mandated testing issues adequately reflects the complexity of teaching. (Item 38)
- Teachers in my school have found ways to raise state-mandated test scores without really improving student learning. (Item 45)
- The state-mandated test is not an accurate measure of what students who are acquiring English as a second language know and can do. (Item 31)
- Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness. (Item 25)

In addition to the items on the scale, several other germane survey items will be discussed at the item level. Most items that make up the value scale were coded 1 for "strongly disagree," 2 for "disagree", 3 for "agree" and 4 for "strongly agree"; higher values for individual items represent greater agreement or the perception that the state test was valuable. Negatively worded items were coded in reverse order to maintain a common interpretation of the scale. Scale scores were computed by averaging responses to the 13 survey items. A two-way analysis of variance — stakes (H/H, H/M, H/L, M/H, M/L) by school type (elementary, middle, high) — was conducted to determine whether mean differences were statistically significant. The results are presented in Appendix E, Table E8. Neither the main effects for stakes level and school type nor the interaction effect were significant at alpha = .001. In other words, teachers' perceptions of the value of the state test did not depend on the stakes attached to the test or the type of school; their regard for the test is fundamentally similar across stakes and grade levels. The overall mean on the scale was 1.99, placing teachers at the "disagree" point. This indicates that in general teachers do not highly value the state test. Item-level results by stakes level and grade level, however, show some variation.

Item-Level Results by Stakes Level

In order to explore further teachers' regard for the value of the state test, this section discusses individual survey items related to this issue. A substantial minority of teachers, roughly 40%, across the different types of testing programs indicated that the state test had brought much-needed attention to education issues in their state (see Table 11). More teachers in states with high stakes for students (43%) held this view than their counterparts in M/L states (31%). In addition, roughly 50% of teachers in each type of testing program reported that the state test measures high standards of achievement.

Table 11. Value of the State Test: Percent Agreement by Stakes Level[1,2]

Value of the State Test Items	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
The state-mandated test has brought much-needed attention to education issues in my district.	43	35	38	43	31
The state-mandated test measures high standards of achievement.	48	56	53	45	52
Overall, the benefits of the state-mandated testing program are worth the investment of time and money.	30	23	24	22	28
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	40	40	35	40	36
The state-mandated test motivates previously unmotivated students to learn.	9	3	5	7	4
The state-mandated testing program is just another fad.	47	47	50	55	42

1. Shaded values indicate statistically significant percent differences from the moderate/low category (alpha = .001).
2. The strongly agree and agree response categories were collapsed into general-agreement responses.

The results also suggest that teachers question whether these benefits outweigh the costs associated with the test. For example, approximately three-fourths of teachers at each stakes level disagreed that the benefits are worth the investment of time and money. Even in states with little high-stakes accountability (M/L), a large majority of teachers (72%) indicated that the costs outweigh any apparent gains. Few teachers, less than 10% at each stakes level, agreed that the state test motivates previously unmotivated students to learn. With the exception of teachers from M/L testing programs (42%), roughly 1 out of every 2 teachers at the four remaining stakes levels reported that their state testing program was just another fad. These results suggest that a greater percentage of teachers in states with high stakes for students than in M/L stakes states do not view their state testing policy as sustainable.

Another factor that seemed to influence the value teachers place on the state test is their low opinion of its accuracy as measure of student achievement (see Table 12). Few teachers in each of the five types of testing programs regarded the test as an accurate measure of student achievement and educational quality. For example, only about 15% of teachers at each stakes level indicated that scores on the state test accurately reflected the quality of education students had received. Likewise, 15% across stakes levels agreed with the statement, "The state-mandated test is as accurate a measure of students' achievement as a teacher's judgment." These data clearly show that a large majority of teachers feel that their state test is not indicative of educational quality. Further, teachers question the capacity of the test to accurately measure the achievement of specific student populations. An overwhelming proportion of teachers, approximately 95% in each type of testing program, maintained that the state test does not accurately measure what students who are acquiring English as a second language (ESL) know and can do. A slightly smaller yet sizable percent of teachers across stakes levels, roughly 75%, reported the same view for minority students. Roughly 85% of teachers across stakes levels felt that score differences from year to year reflect changes in the characteristics of students rather than in school effectiveness. Teachers reported the same view about differences in test performance among schools.

Table 12. Test as a Measure of Achievement: Percent Agreement by Stakes Level[1,2]

Student Achievement Items	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Scores on the state-mandated test results accurately reflect the quality of education students have received.	20	16	15	10	19
The state-mandated test is as accurate a measure of student achievement as a teacher's judgment.	19	15	16	12	17
The state-mandated test is NOT an accurate measure of what students who are acquiring English as a second language know and can do.	94	92	92	95	94
The state-mandated test is NOT an accurate measure of what minority students know and can do.	76	74	74	77	72
Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness.	81	86	84	84	86
Differences among schools on the state-mandated test are more a reflection of students' background characteristics than of school effectiveness.	85	75	86	88	84
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	40	40	35	40	36
Performance differences between minority and non-minority students are smaller on the state-mandated test than on commercially available standardized achievement tests (e.g. Stanford 9, ITBS, CAT).	23	17	21	18	27

1. Shaded values indicate statistically significant percent differences from the moderate/low category (alpha = .001).
2. The strongly agree and agree response categories were collapsed into general-agreement responses.

It may be that teachers' perceptions of the value of the state test are mitigated by their practices to improve student performance. Over one-third of teachers in each type of testing program indicated that teachers in their school have found ways to raise state-mandated test scores without really improving student learning. In addition, roughly 20% across stakes levels indicated that performance differences between minority and non-minority students were smaller on the state test than on other commercial standardized achievement tests. However, fewer teachers in H/M (17%) and M/H stakes states (18%) than in M/L stakes states (27%) so reported.

External factors such as the media's reporting of test results may also influence teachers' opinions of the value of the state test. In general, teachers viewed test-related media coverage negatively (see Table 13). For example, roughly 90% at each stakes level disagreed that "the media coverage of state-mandated test results accurately depicts the quality of education in my state." Almost 9 out of 10 teachers, across all types of testing programs indicated that media coverage of state-mandated testing issues has been unfair to teachers. Only a small percentage (roughly 10%) across stakes levels reported that media coverage adequately reflects the complexity of teaching. At the very least, a substantial proportion of teachers suggest that when test results are reported without recognizing the realities of teaching and the context in which schools and classrooms operate, their perceptions of the value of the state test may be negatively affected.

Table 13. Media Coverage: Percent Agreement by Stakes Level[1,2]

Media-Related Items	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Media coverage of state-mandated test results accurately reflects the quality of education in my state.	14	12	8	7	11
Media coverage of state-mandated testing issues has been unfair to teachers.	86	88	87	89	84
Media coverage of state-mandated testing issues adequately reflects the complexity of teaching.	14	11	10	10	10

1. Shaded values indicate statistically significant percent differences from the moderate/low category (alpha = .001).
2. The strongly agree and agree response categories were collapsed into general-agreement responses.

Item-Level Results by School Type

Item-level results by school type (elementary, middle, high) show some variation in teachers' perceptions of the value of the state test. Teachers' opinions across stakes levels were fairly consistent concerning their belief that the test measures high standards of achievement (see Table 14). When the same item is examined according to grade level, an interesting pattern emerges. As the grade level increases, the proportion of teachers who reported that their state test measured high standards of achievement decreased. More elementary (56%) and middle school teachers (48%) held this view than did high school teachers (35%). However, teachers across grade levels are in general agreement about the power of the state test to draw attention to education issues and other benefits of the test. Roughly 40% at each grade level agreed that "the state-mandated test has brought much-needed attention to education issues in my district." However, roughly 70% of teachers at each grade level disagreed that the benefits of the program are worth the investment of time and money. A significantly larger percentage of high school (44%) than elementary school teachers (38%) reported that teachers in their school had found ways to raise test scores without improving student learning.

Table 14. Value of the State Test: Percent Agreement by School Type[1,2,3]

Value of the State Test Items	School Type		
	Elementary	Middle	High
The state-mandated test has brought much-needed attention to education issues in my district.	43	38	37
The state-mandated test measures high standards of achievement.	*56*	48	35
Overall, the benefits of the state-mandated testing program are worth the investment of time and money.	27	32	27
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	38	40	44
The state-mandated test motivates previously unmotivated students to learn.	6	9	10
The state-mandated testing program is just another fad.	47	47	52

1. Shaded values indicate statistically significant percent differences from the high school category (alpha = .001).
2. Italicized values indicate statistically significant percent differences between the elementary and middle school results at alpha = .001.
3. The strongly agree and agree response categories were collapsed into general-agreement responses.

BEST COPY AVAILABLE

While results suggest that high school teachers place less value on the state test than teachers in lower grades, elementary teachers showed a greater concern for the accuracy of the test as an indicator of school effectiveness (see Table 15). Greater percentages of elementary (83%) and middle (85%) than high school teachers (77%) reported that score differences from year to year reflect changes in the characteristics of students rather than in school effectiveness. In addition, more elementary teachers indicated that the state test is not an accurate measure of achievement for ESL and minority students, even while a substantial majority of teachers at each grade level agreed with the statement. Ninety-five percent of elementary teachers and 90% of high school teachers indicated that the test inaccurately measures student achievement for English as a second language students. Similarly, slightly more elementary (78%) than high school teachers (71%) reported that the state test was not an accurate measure of minority students' achievement. Roughly 85% of teachers across grade levels attributed differences in performance among schools to student characteristics rather than school effectiveness.

Table 15. Test as a Measure of Achievement: Percent Agreement by School Type[1,2,3]

Student Achievement Items	School Type		
	Elementary	Middle	High
Scores on the state-mandated test results accurately reflect the quality of education students have received.	18	17	18
The state-mandated test is as accurate a measure of student achievement as a teacher's judgment.	17	18	17
The state-mandated test is NOT an accurate measure of what students who are acquiring English as a second language know and can do.	95	94	90
The state-mandated test is NOT an accurate measure of what minority students know and can do.	***78***	72	71
Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness.	83	85	77
Differences among schools on the state-mandated test are more a reflection of students' background characteristics than of school effectiveness.	85	88	83
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	38	40	44
Performance differences between minority and non-minority students are smaller on the state-mandated test than on commercially available standardized achievement tests (e.g. Stanford 9, ITBS, CAT).	21	22	20

1. Shaded values indicate statistically significant percent differences from the high school category (alpha = .001).

2. Italicized values indicate statistically significant percent differences between the elementary and middle school results at alpha = .001.

3. The strongly agree and agree response categories were collapsed into general-agreement responses.

Just as teachers' opinions of media coverage of education issues and the state test were fairly consistent across types of testing programs, responses varied only slightly by grade level (see Table 16). Generally, a substantial majority of teachers across grade levels reported negative opinions about media coverage of the state test. A substantial majority disagreed with the statement that "media coverage of the state-mandated test results accurately reflects the quality of education in my state": 86% of elementary, 88% of middle, and 91% of high school teachers. Similarly, an overwhelming majority of teachers in each type of school reported that media coverage of testing issues has been unfair to teachers: 89% of elementary teachers and 84% of both middle and high school teachers. More than 85% of elementary, middle and high school teachers reported that media coverage of testing issues did not adequately reflect the complexities of teaching.

Table 16. Media Coverage: Percent Agreement by School Type[1,2,3]

Media-Related Items	School Type		
	Elementary	Middle	High
Media coverage of state-mandated test results accurately reflects the quality of education in my state.	14	12	9
Media coverage of state-mandated testing issues has been unfair to teachers.	**89**	84	84
Media coverage of state-mandated testing issues adequately reflects the complexity of teaching.	13	12	10

1. *Shaded values indicate statistically significant percent differences from the high school category (alpha = .001).*
2. *Italicized values indicate statistically significant percent differences between the elementary and middle school results at alpha = .001.*
3. *The strongly agree and agree response categories were collapsed into general-agreement responses.*

Summary

Standards-based reform was implemented in response to the demand for higher standards and increased student achievement. Even though a substantial proportion of teachers recognized that state testing programs have refocused attention on important educational issues and reflect high academic standards, they place less value on the state test as an accurate measure of student achievement or as an indicator of educational quality. Generally, teachers' views on the value of the state test are highly negative and fairly consistent regardless of the type of testing program and school in which they work. In addition, the survey results show that teachers' feel ill-used by the media, which they feel does not understand the complexity of teaching or the many factors affecting learning.

V. Impact of the State Test on Content and Mode of Instruction

The assumption underpinning the establishment of standards and test-based accountability systems is that they motivate teachers and schools to improve student learning and focus on specific types of learning. Some observers have raised concerns that the latter too often translates into "teaching to the test." As Shepard (1990) notes, however, teaching to the test means different things to different people. Some state and local educational leaders, as well as classroom teachers, interpret the phrase to mean "teaching to the domain of knowledge represented by the test" (p. 17) rather than narrowly teaching only the content and items expected to be on the test. By this definition, many would argue that one goal of testing is to influence what teachers teach. After interviews with state testing directors in 40 high-stakes states, Shepard writes:

> *When asked, "Do you think that teachers spend more time teaching the specific objectives on the test(s) than they would if the tests were not required?" the answer from the 40 high-stakes states was nearly unanimously, "Yes." The majority of respondents [described] the positive aspects of this more focused instruction. 'Surely there is some influence of the content of the test on instruction. That's the intentional and good part of testing, probably.'... Other respondents (representing about one third of the high-stakes tests) also said that teachers were spending more time teaching the specific objectives on the test but cast their answer in a negative way: 'Yes.... There are some real potential problems there.... Basically the tests do drive the curriculum.' (p. 18)*

In the remainder of this section, we focus on survey items, which asked teachers whether and how the content and mode of instructional practices are being influenced by the state-mandated test. This discussion is based on teachers' responses to two survey items (item 62 and 76); each was composed of several additional items. Item 62 presented teachers with various content areas and asked, "In what ways, if any, has the amount of time spent on each of the following activities changed in your school in order to prepare students for the state-mandated testing program?" Teachers selected from five response options ranging from "decreased a great deal" (1) to "increased a great deal" (5). While Item 62 dealt with content areas, Item 76 dealt with methods of instruction. It asked teachers to indicate the extent to which they agreed with the statement: "Your state-mandated testing program influenced the amount of time you spend on..." followed by a number of pedagogical practices or instructional emphases (e.g. whole-group instruction, critical thinking skills, individual-seat work).

Impact on Instructional Content and Activities

Using factor analytic techniques (see Appendix E, Table E9), the items composing question 62 were combined to form three scales: (1) Impact on Tested Subject Areas, (2) Impact on Non-Core Subject Areas, and (3) Impact on Student and Class Activities. Table 17 presents the items that each scale comprises from Item 62.

Table 17. Items Comprised by the Tested Areas, Non-Core Content, and Classroom Activities Scales

Item 62: In what ways, if any, has the amount of time you spent on each of the following activities changed in your school in order to prepare students for the state-mandated testing program?	Scale
Instruction in tested areas	Tested Areas
Instruction in areas not covered by the state-mandated test	Tested Areas
Instruction in tested areas with high stakes attached (e.g., promotion, graduation, teacher rewards)	Tested Areas
Parental contact	Tested Areas
Instruction in fine arts	Non-Core
Instruction in physical education	Non-Core
Instruction in foreign language	Non-Core
Instruction in industrial/vocational education	Non-Core
Student free time (e.g., recess, lunch)	Activities
Field trips (e.g., museum tour, hospital tour)	Activities
Class trips (e.g., circus, amusement park)	Activities
Student choice time (e.g., games, computer work)	Activities
Organized play (e.g., games with other classes)	Activities
Enrichment school assemblies (e.g., professional choral group performances)	Activities
Administrative school assemblies (e.g., awards ceremonies)	Activities
Classroom enrichment activities (e.g., guest speakers)	Activities
Student performance (e.g., class plays)	Activities

The three scales were used to compare the impact of testing across the five types of state testing programs. Two-way analyses of variance — stakes level by school type — were conducted to determine whether mean differences on the three scales were statistically significant. For each scale both the main effect for stakes level and school type were significant at alpha = .001, however the interaction effect was not (see Appendix E, Tables E10-E12). Table 18 displays the mean scale scores for each program. For each scale, lower mean values represent decreased time and higher mean values represent increased time. For all state testing programs, teachers indicated that they have increased instruction in tested areas. The largest increases occurred in H/H and M/H programs, and the smallest in H/M and M/L programs.

Table 18. Means on the Tested Areas, Non-Core Content, and Classroom Activities Scales by Stakes Levels

Scales	Stakes Level Scale Mean				
	H/H	H/M	H/L	M/H	M/L
Tested-Areas	3.85	3.53	3.65	3.80	3.47
Non-Core Areas	2.66	2.85	3.01	2.74	2.87
Activities	2.47	2.73	2.68	2.55	2.73

Table 18 also indicates that teachers reported decreased time spent on activities and non-core subject areas across all testing programs. In contrast to instruction in tested areas, time spent on activities and non-core subject areas decreased the most in H/H and M/H programs and the least in H/M and M/L programs.

Table 19 displays the mean scale scores for elementary, middle, and high school teachers. For all school levels, instruction in tested areas increased while time spent on activities and non-core subject areas decreased. Although the impact was similar at all three levels, the smallest increases and decreases occurred at the high school level. Note that all differences between elementary and high school were statistically significant at the .001 level. Only the difference between the middle and high school level for tested areas was statistically significant.

Table 19. Means on the Tested Areas, Non-Core Content, and Classroom Activities Scales by School Type

Scales	School Type Scale Mean		
	Elementary	Middle	High
Tested Areas	3.73	3.68	3.54
Non-Core Areas	2.79	2.81	2.88
Activities	2.60	2.61	2.71

In general, Table 19 shows time spent on tested areas increased the most in elementary and middle schools; these increases were the largest in H/H and M/H states. The largest decrease in time spent on activities and non-core areas generally occurs in elementary and middle school and in the H/H and M/H states. In summary, the data suggest that instructional practices are affected by testing programs.

Impact on Instruction in Tested and Non-Tested Areas: Item-Level Results

As at the scale level, the perceived impact on instructional practices is strongest for teachers in H/H states and weakest for those in M/L states when items are examined individually. As shown in Table 20, 43% of teachers in H/H states indicated that instruction in tested areas has increased a great deal. In contrast, only 17% of teachers in M/L states so reported. Between 32% and 35% of teachers in H/M, H/L, and M/H states indicated that their instruction in tested areas increased greatly. The data also show that 32% of teachers in M/L states indicated that their instruction in tested areas had not changed, as compared with 20% of teachers in H/H states and 17% in M/H states.

Table 20. Tested and Non-Tested Content Areas: Percent Reporting Change in Instructional Time[1,2]

Change in time spent on instruction in:	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Tested areas					
Decreased a great deal	0	0	0	0	0
Moderately decreased	1	2	1	1	0
Stayed about the same	20	22	23	17	32
Moderately increased	36	44	43	46	51
Increased a great deal	43	32	34	35	17
Areas not covered by the state test					
Decreased a great deal	25	14	19	23	9
Moderately decreased	34	28	36	40	33
Stayed about the same	36	48	38	31	51
Moderately increased	4	7	5	5	7
Increased a great deal	1	3	2	1	1
Tested areas with high-stakes attached					
Decreased a great deal	1	2	1	0	2
Moderately decreased	1	2	2	1	2
Stayed about the same	37	64	56	40	66
Moderately increased	34	23	27	33	20
Increased a great deal	27	10	14	26	10

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

BEST COPY AVAILABLE

For instruction in non-tested areas, Table 20 shows that teachers in both H/H and M/H states indicated the greatest decreases (25% and 23%, respectively). While such instruction has also decreased in the other states, only 9% of teachers in M/L states indicated great decreases while 51% of them indicated no change. In addition, instruction in areas with high stakes attached increased in all testing programs. The largest increases occurred in H/H and M/H states and the smallest in M/L and H/M states.

Impact on Non-Core Subjects: Item-Level Results

Table 21 presents the results for three items that form the Non-Core Subject Area scale. Most teachers in all states indicated that instruction in fine arts has remained the same. A higher percentage of teachers in H/H states indicated that instruction in fine arts has decreased greatly. About the same percentage of teachers in H/H, M/H, and M/L states indicated moderate decreases. Interestingly, teachers in H/L states indicated the largest increases (both great and moderate). These are likely due to testing in the area of art in three of the H/L states (Kentucky, Missouri, Oklahoma).

Table 21. Non-Core Content Areas: Percent Reporting Change in Instructional Time[1,2]

Change in time spent on instruction in:	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Fine Arts					
Decreased a great deal	16	8	6	12	7
Moderately decreased	19	12	11	21	18
Stayed about the same	60	70	55	63	64
Moderately increased	5	9	21	4	10
Increased a great deal	1	1	7	0	1
Physical Education					
Decreased a great deal	9	3	3	4	3
Moderately decreased	15	12	8	13	14
Stayed about the same	74	82	78	81	79
Moderately increased	2	5	9	2	4
Increased a great deal	0	1	2	0	1
Foreign Language					
Decreased a great deal	11	6	6	7	7
Moderately decreased	10	8	6	12	8
Stayed about the same	70	80	75	78	76
Moderately increased	7	6	10	4	10
Increased a great deal	1	1	2	0	1
Industrial/Vocational Education					
Decreased a great deal	16	9	6	9	6
Moderately decreased	15	9	9	15	11
Stayed about the same	64	78	76	73	76
Moderately increased	4	5	9	2	7
Increased a great deal	1	0	1	0	1

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Most teachers within each type of testing program also indicated that instruction has remained about the same in physical education. As in fine arts, teachers in H/H, M/H and M/L reported the largest decreases in instruction in physical education while teachers in H/L states reported the largest increases. Again, this increase seems to be related to physical education standards and tests in two of the states in the H/L group (Missouri and Rhode Island). Similar patterns emerged when teachers responded on the amount of time devoted to industrial/vocational education in preparing students for the state test; most teachers across all states reported that the time has remained the same. The largest decreases in instructional time occurred in H/H and M/H states, with 31% of H/H and 24% of M/H teachers reporting decreased instructional time. Very few teachers reported that instruction in this area had increased.

Impact on Classroom Activities: Item-Level Results

Table 22 displays results for four of the items that form the Activities scale. Across all four items, more teachers in H/H and M/H testing programs indicated that time spent on these activities has decreased. As the data show, most teachers in all testing programs indicated that field trips have been largely unaffected by state testing programs. Compared with the other testing programs, however, fewer teachers in H/H states (60%) so indicated. Teachers in H/H (24%), M/H (22%), and M/L states (22%) were the most likely to report moderate decreases. H/H and M/H states also contain the highest percentage of teachers who reported great decreases in field trips (14% and 11% respectively). The largest decreases in organized play were reported by teachers in H/H and M/H states, where 55% and 46%, respectively, reported spending less time on activities such as structured games with other classes. The H/M and M/L states had the greatest percentage of teachers who indicated that time allocated for organized play has remained the same.

Similar patterns emerged when teachers were asked about the time devoted to class enrichment activities. The largest decreases in activities such as having guest speakers were in the H/H and M/H states. Roughly, a third of teachers in each of the testing programs (34% of H/H and 33% of M/H) reported that they spent less time on enrichment activities so that they could prepare students for the state test. Teachers who reported that the time stayed about the same were typically from H/M and M/L states. Teachers' responses also suggest that student performances have been largely unaffected by state testing programs, particularly in H/M and M/L states. The largest negative impact occurred in the H/H and M/H states. Although very few teachers in M/L states reported a great decrease in student performances, a small but substantial percentage (19%) indicated moderate decreases. Few teachers in any state report increases in student performances in response to state testing programs.

Table 22. Classroom Activities: Percent Reporting Change in Instructional Time[1,2]

Change in time spent on instruction in:	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Field Trips					
Decreased a great deal	14	8	10	11	5
Moderately decreased	24	12	18	22	22
Stayed about the same	60	75	62	65	71
Moderately increased	2	5	8	3	3
Increased a great deal	0	0	1	0	0
Organized Play					
Decreased a great deal	26	14	16	22	9
Moderately decreased	29	20	23	24	24
Stayed about the same	45	62	59	53	65
Moderately increased	1	3	3	1	2
Increased a great deal	0	0	0	0	0
Class Enrichment Activities					
Decreased a great deal	13	5	8	9	5
Moderately decreased	21	14	17	24	15
Stayed about the same	60	71	63	61	72
Moderately increased	6	9	12	6	8
Increased a great deal	0	1	1	0	0
Student Performance					
Decreased a great deal	19	8	12	12	5
Moderately decreased	19	15	16	23	19
Stayed about the same	57	70	64	63	70
Moderately increased	4	7	8	2	6
Increased a great deal	1	1	1	0	0

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Pedagogy and Instructional Emphasis

The survey contained seven items that focused on teaching practice. For each item, teachers were asked to indicate the extent to which they agreed with the following statement: "Your state-mandated testing program influences the amount of time you spend on…" followed by a specific pedagogical practice or instructional emphasis. When examining the findings for these seven items, note that agreement indicates only that a given practice has been affected by the state testing program; this effect, however, could be either positive or negative.

Across all seven items, teachers generally agreed that the state test has influenced their use of specific pedagogical practices and instructional emphases. For all items, significantly more teachers in H/H programs strongly agreed on this effect. Conversely, across most items, significantly fewer teachers in M/L programs strongly agreed. As Table 23 shows, a higher percentage of teachers in H/H programs (72%) agreed that the state testing program was influencing their whole-group instruction while only 51% of teachers in M/L programs agreed. This opinion was particularly acute for H/H teachers, of whom 26% strongly agreed in comparison with 8% of M/L teachers. Conversely, 40% of M/L teachers disagreed that their use of whole-group instruction had been influenced by the testing program, as compared with 23% in H/H programs. Teachers in the three other programs generally fell between these two extremes.

Table 23. Methods of Instruction: Percent Agreement by Stakes Level[1,2]

Your state testing program has influenced the amount of time you spend on:	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Whole-group instruction	72	59	69	64	51
Critical thinking skills	81	72	79	75	63
Individual seat work	64	51	59	55	43
Basic skills	83	72	77	74	68
Cooperative learning	60	52	58	54	41
Concept development	65	56	62	55	48
Problems likely to appear on test	81	71	75	73	59

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Teachers generally agreed that their state testing program had also influenced their focus on students' critical thinking skills. Many more teachers in H/H programs (81%) agreed that their focus had been influenced than did M/L teachers (63%). Again, the views held by H/H teachers were particularly intense compared with those of their M/L counterparts: almost three times as many teachers in H/H states (30%) as in M/L states (12%) strongly agreed that the time allocated to critical thinking skills was influenced by the state test.

With regard to individual-seat work, many more H/H teachers (64%) agreed that instruction had been influenced than M/L teachers (43%). Similar patterns related to the intensity of teachers' opinions also emerged. For example, H/H teachers were more apt than M/L teachers to strongly agree with the statement (19% and 7% respectively). Conversely, considerably fewer H/H teachers disagreed and strongly disagreed (32% and 5%) that individual-seat work had been influenced, while a significantly higher percentage of M/L teachers disagreed (57%). Teachers in other programs generally fell between these two extremes.

The largest differences in the influence of testing programs on the amount of time teachers spend on basic skills occurs between the H/H and M/L programs. Across all five testing programs, most teachers agreed that the program influenced the time spent on basic skills. Many more teachers in H/H programs, however, strongly agreed, versus those in M/L programs (30% and 13% respectively). Conversely, a higher percentage of M/L teachers disagreed (32%) that time spent on basic skills has been influenced, compared with 17% of teachers in H/H programs.

In general, state testing programs appear to have had less influence on the time teachers spend on cooperative learning activities. As with other instructional practices, many teachers in H/H settings strongly agreed that cooperative learning has been influenced by testing while fewer of them disagreed; conversely, fewer M/L teachers agreed or strongly agreed and more of them disagreed. Teachers in the three other programs generally fell between these two extremes. The data show that between 48% and 65% of teachers in all testing programs agreed that testing has influenced the time spent on concept development through the use of manipulatives or experiments. Greater percentages of teachers in H/H settings strongly agreed (20%) while greater percentages in M/L programs disagreed (52%) that this is the case.

A majority of teachers in all settings agree that testing has influenced the time spent on problems likely to appear on the test. Noticeably fewer teachers in M/L programs strongly agreed that the time they spend on such problems has been affected (11%) and more teachers in H/H settings did so (32%) than their counterparts in other programs.

Summary

Based on teachers' responses to the survey items examined in this section, it appears that state testing programs are influencing both what teachers teach and how they teach. Across all types of testing programs, teachers reported increased time spent on subject areas that are tested and less time on those that are not. In addition, teachers in all programs reported that testing has influenced the amount of time spent on activities not directly related to specific subject areas. Similarly, the majority of teachers in all testing programs agreed that the state testing programs are influencing the amount of time they spend using a variety of instructional methods, such as whole-group instruction, individual-seat work, cooperative learning, and using problems similar to those on the test. In general, the influence of state testing programs on teachers' instructional practices is stronger in H/H settings than in M/L settings. Moreover, testing programs appear to be having a strong influence on the amount of time teachers in M/H settings spend on tested areas and on activities. Thus, it appears that the influence on the subject areas tested is more closely related to the stakes for students than to those for schools. Finally, the impact of testing programs is generally stronger in elementary and middle schools than in high schools.

VI. The Impact of the State Test on Preparation Practices

Historically, test preparation is a persistent issue associated with high-stakes testing programs. Teachers have always tried through various means to prepare students for the hurdle of a high-stakes test. When the stakes are also high for teachers, schools, or districts there is an added incentive to have students perform well on state-mandated tests. To answer the question, "What do teachers do to prepare students for state-mandated tests?" the survey included a section on test preparation practices. Teachers responded to the following seven items:

- Item 60 asked how teachers prepare students for the test.
- Item 63 asked about the number of class hours per year of test preparation.
- Item 64 asked when test preparation begins.
- Item 65 asked about the similarity of the test preparation content to the test itself.
- Item 66 asked whether teachers targeted various groups of students for preparation.
- Item 67 asked whether teachers had heard about various activities by other teachers during the test administration.
- Item 68 asked about the use of ways of motivating students to do their best on the test. (For the exact wording of these items see Appendix A.)

Test Preparation Practices

Table 24 examines teachers' responses to Item 60 across the five stakes levels (H/H, H/M, H/L, M/H, and M/L). Teachers indicated whether or not they used each of eight approaches to test preparation.

Examination of Table 24 reveals significant differences among the five stakes levels on each of the eight practices. Fewer teachers in the H/H states chose the option "I do no preparation" than teachers in the M/L states. The converse was true for the other seven practices listed; i.e., more teachers in the H/H states and fewer in the M/L states chose these options.

The table shows that the largest differences between the H/H and the M/L teachers are around the last three practices: "I provide students with items similar to those on the test" (75% vs. 54%); "I provide test-specific preparation materials developed commercially or by the state" (63% vs. 19%); and "I provide students with released items from the state-mandated test" (44% vs. 19%).

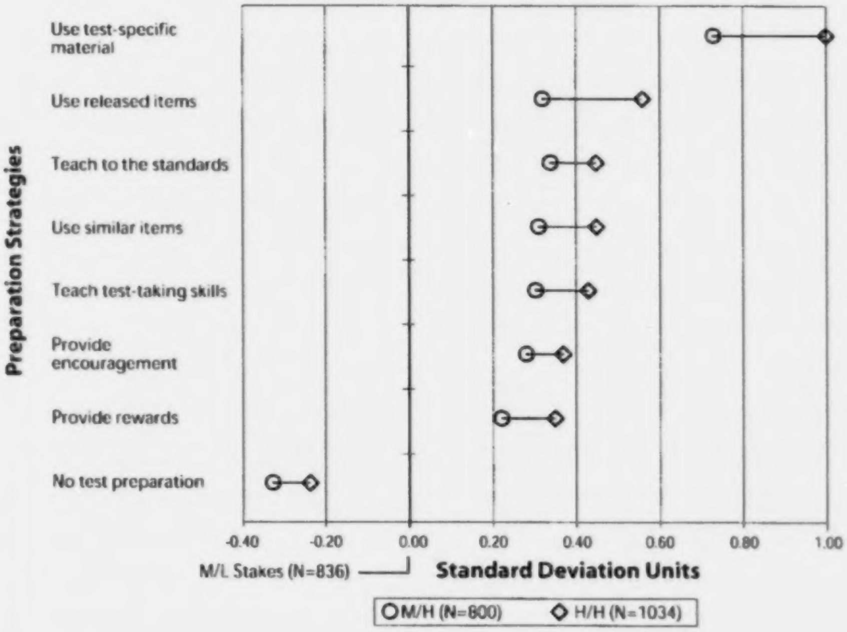
Table 24. Test Preparation Strategies: Percent Reporting by Stakes Level[1,2]

Test Preparation Strategies	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
I do no special test preparation.	10	14	12	13	22
I teach test taking skills.	85	72	76	80	67
I encourage students to work hard and prepare.	83	78	78	79	67
I provide rewards for test completion.	20	14	15	15	8
I teach the standards or frameworks known to be on the test.	75	62	69	70	54
I provide students with items similar to those on the test.	75	65	73	69	54
I provide test-specific preparation materials developed commercially or by the state.	63	47	45	52	19
I provide students with released items from the state-mandated test.	44	30	47	33	19

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Finally, Figure 6 shows the difference in the preparation practices of teachers in the states with high stakes for students (H/H and M/H) and those with M/L stakes testing programs. For comparison purposes, the proportion of M/L teachers who indicated using each practice listed in Item 60 was set to zero. Then the distance of the H/H and M/H groups' proportions from that of the M/L group on each practice was plotted in terms of standard deviations units. The differences are between .2 and .6 of a standard deviation away from the M/L group except for the practice "I provide test-specific preparation materials developed commercially or by the state." There the distance of M/H and H/H proportions jumps to almost .8 to a full standard deviation away from that of M/L teachers.

Figure 6.
Use of Test Preparation Strategies: H/H and M/H vs. M/L Stakes States



Another way to consider teachers' responses to Item 60 is to cross the five stakes levels by teachers' grade level — elementary, middle, and secondary. Table 25 shows the percentages for the 8 practices by each grade in Item 60.

For the practice "I do no preparation," Table 25 shows that fewer elementary teachers in the H/H states and more in the M/L states chose this reply. The percentages of these teachers, however, are relatively small — 6% of H/H and 18% of M/L elementary teachers. As for the practice "I provide rewards for test completion," more elementary teachers in the H/H states (25%) and fewer in the M/L states (10%) indicated using this technique.

The practice "I provide test-specific preparation materials developed commercially or by the state" shows a large difference between teachers in H/H and in M/L states regardless of grade level. Seventy-one percent of H/H elementary teachers chose this practice, compared with 24% of M/L teachers. The same pattern holds in middle school (H/H 60% vs. M/L 13%) and in high school (H/H 46% vs. M/L 11%). Likewise, the practice "I provide students with released items from the state-mandated test" was chosen by considerably more teachers in the H/H states than by their counterparts in the M/L states regardless of grade level. Close to 45% of H/H teachers across the three grade levels indicated that they used released items, compared with 16 to 24% of M/L teachers.

Table 25. Test Preparation Strategies: Percent Reporting by Stakes Level and School Type[1,2]

Test Preparation Strategies	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
I do no special test preparation.	Elementary	6	10	11	10	18
	Middle	10	16	9	12	25
	High	19	23	17	21	31
I teach test taking skills.	Elementary	90	80	82	87	72
	Middle	83	64	75	78	64
	High	71	58	59	63	54
I encourage students to work hard and prepare.	Elementary	83	82	80	81	71
	Middle	87	75	80	77	67
	High	77	68	69	76	58
I provide rewards for test completion.	Elementary	25	19	16	20	10
	Middle	15	11	19	8	2
	High	12	5	10	9	7
I teach the standards or frameworks known to be on the test.	Elementary	75	66	74	75	57
	Middle	79	62	65	70	52
	High	71	55	60	59	47
I provide students with items similar to those on the test.	Elementary	78	70	78	72	58
	Middle	72	64	72	71	53
	High	68	54	59	58	44
I provide test-specific preparation materials developed commercially or by the state.	Elementary	71	56	55	60	24
	Middle	60	44	42	50	13
	High	46	29	21	33	11
I provide students with released items from the state-mandated test.	Elementary	44	28	52	34	19
	Middle	46	35	51	31	24
	High	42	31	32	32	16

1. Overall chi-square for each item is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

It is clear from responses to Item 60 that attaching high stakes to test performance encourages many teachers to use various test preparation tactics to improve their students' performance. The data show that the practices of teachers in the high-stakes states differ significantly from those of their counterparts in states where the stakes are not as high. This is particularly true for two practices: "I provide test-specific preparation materials developed commercially or by the state," and "I provide students with released items from the state-mandated test." When high stakes are attached to scores, test preparation becomes more test-specific. The danger with this is that historically test-specific practices, such as teaching to past test questions, can corrupt the validity of the test itself.

Class Hours Per Year Spent on Test Preparation

Table 26 shows teachers' responses to Item 63, "Approximately how many class hours PER YEAR do you spend preparing students specifically for the state-mandated test (i.e., teaching test taking skills)?" Seventeen percent of teachers from M/L states chose "None" compared with only 5% of those from H/H states. Further, 51% of M/L teachers chose the "1-10 hours" response, compared with only 24% of H/H teachers. The largest difference can be seen in the "more than 30 hours" option, the choice of 44% of H/H teachers vs. only 10% of M/L teachers.

Table 26. Test Preparation Time: Percent Reporting by Stakes Level[1,2]

Class hours per year spent	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
None	5	11	7	9	17
1-10 hours	24	33	32	33	51
11-20 hours	14	18	20	19	15
21-30 hours	13	9	11	9	7
More than 30 hours	44	28	30	31	10

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Figure 7.
Test Preparation Hours: H/H and M/H vs. M/L Stakes States

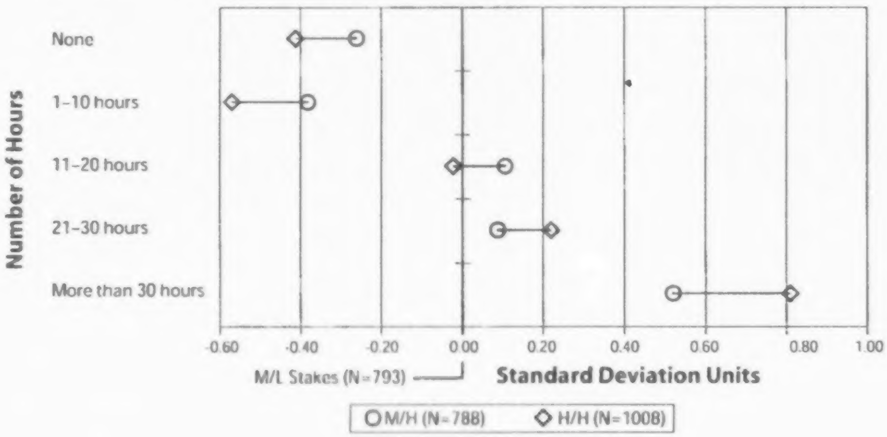


Figure 7 shows how the proportions of H/H and M/H teachers differ from that of the M/L teachers for time spent on test preparation. The "more than 30 hours" option for the H/H teachers is .8 of a standard deviation away from the M/L group while the M/H teachers are more than half of a standard deviation away.

Table 27 shows the percentages associated with each response option for the number of hours teachers devote to test preparation (Item 63) by stakes level and grade. The number of hours devoted to test preparation is higher for all teachers in the high-stakes states (H/H, H/M, H/L, M/H) than in the M/L states. The highest percentage of elementary teachers in the high-stakes states choosing the option "more than 30 hours" ranged from 51% of H/H elementary teachers to over 36% of H/M, H/L and M/H teachers. This compares with 12% of elementary teachers in the M/L states. Middle school teachers exhibited the same response pattern for the "more than 30 hours" option, but the percentages are lower 42% (H/H), 20% (H/M), 29% (H/L) and 27% (M/H). Only 7% of middle school teachers in M/L states chose the "more than 30 hours" option.

Table 27. Test Preparation Time: Percent Reporting by Stakes Level and School Type[1,2]

Class hours per year spent	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
None	Elementary	3	6	6	6	12
	Middle	3	16	3	7	20
	High	12	19	17	18	28
1-10 hours	Elementary	20	26	27	28	49
	Middle	27	36	34	37	53
	High	34	48	43	40	53
11-20 hours	Elementary	12	18	21	20	18
	Middle	21	22	21	20	16
	High	16	17	18	16	7
21-30 hours	Elementary	15	11	11	10	9
	Middle	7	7	13	10	5
	High	13	7	9	7	3
More than 30 hours	Elementary	51	38	36	36	12
	Middle	42	20	29	27	7
	High	25	10	13	19	9

1. Overall chi-square for each item is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

It is clear that across stakes levels elementary teachers reported spending more time in test preparation than did secondary school teachers. The percentage of high school teachers choosing the "more than 30 hours" option is considerably lower than that of teachers in the lower grades (25%, 10%, 13%, 19% and 9% of high school teachers across stakes levels). High school teachers were more apt than their counterparts in other grades to choose the 1-10 hour option (34%, 48%, 43%, 40% and 53% across stakes levels). This may be due to the fact that elementary teachers have to prepare their students for a battery of tests covering two to four elementary subjects while secondary teachers specialize in a specific discipline and need to concern themselves with only one test.

These data for time spent on test preparation show that stakes levels associated with a state's testing program strongly influence the amount of time teachers spend on test preparation — the higher the stakes, the more class time was spent on test preparation, particularly at the elementary level.

When Test Preparation Activities Were Conducted

Table 28 shows the responses of teachers to Item 64, "When were most of the test preparation activities you conducted specifically for the state-mandated test carried out?" The largest differences were between the H/H teachers and the M/L teachers. For example, M/L teachers were more apt to select the "no specific preparation" and "throughout the week before" options than were H/H teachers (20% vs. 5% and 10% vs. 4%, respectively).

Table 28. Timing of Test Preparation: Percent Reporting by Stakes Level[1,2]

Timing of Test Preparation	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
No specific preparation	5	11	8	8	20
The day before the state test	1	1	2	2	2
Throughout the week before the state test	4	8	6	6	10
Throughout two weeks before the state test	6	7	8	6	8
Throughout the month before the state test	14	19	15	20	17
Throughout the year	70	54	62	58	43

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Teachers in H/H states are much more likely to prepare "throughout the year" than are teachers from M/L states (70% compared with 43%). The difference on this time option is seen graphically in Figure 8: the proportion of H/H teachers is close to .6 of a standard deviation away from the M/L teachers.

Figure 8.
Test Preparation Timing: H/H and M/H vs. M/L Stakes States

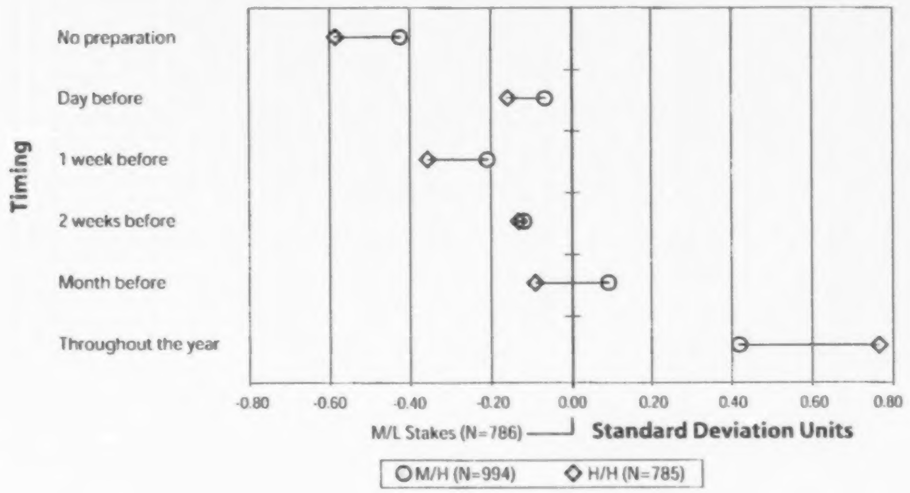


Table 29 shows that for elementary teachers the option chosen by the most teachers is "throughout the year" — 76% (H/H), 62% (H/M), 66% (H/L), 65% (M/H), and 46% (M/L) across stakes levels. The same response pattern holds for middle school teachers — 72% (H/H), 48% (H/M), 62% (H/L), 55% (M/H) and 41% (M/L). For high school teachers the pattern is less pronounced — 53% (H/H), 39% (H/M), 51% (H/L), 44% (M/H), and 37% (M/L).

Table 29. Timing of Test Preparation: Percent Reporting by Stakes Level and School Type[1,2]

Timing of Test Preparation	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
No specific preparation	Elementary	4	5	6	5	15
	Middle	3	15	4	8	22
	High	13	21	15	18	30
The day before the state test	Elementary	1	1	2	1	2
	Middle	0	1	1	3	2
	High	1	1	3	3	3
Throughout the week before the state test	Elementary	1	5	5	4	9
	Middle	5	7	6	10	13
	High	9	15	9	9	14
Throughout the two weeks before the state test	Elementary	5	7	6	5	8
	Middle	5	8	12	6	10
	High	7	8	9	9	6
Throughout the month before the state test	Elementary	14	20	16	21	21
	Middle	14	22	16	18	13
	High	17	16	14	18	10
Throughout the year	Elementary	76	62	66	65	46
	Middle	72	48	62	55	41
	High	53	39	51	44	37

1. Overall chi-square for each item is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Similarity of Test Preparation Content to the Test

Table 30 shows teachers' responses to Item 65, "How similar is the content of the test preparation materials you use to the content of the state-mandated test?" The greatest differences were found for the "very similar" option, selected by 40% of H/H teachers compared with 20% of the M/L teachers, and the "very dissimilar" option, chosen by 2% of H/H teachers compared with 6% of M/L teachers.

Table 30. Content of Test Preparation Material: Percent Reporting by Stakes Level[1,2]

Test Preparation Content	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Very similar to the content of state test	40	29	32	37	20
Somewhat similar to the content of state test	52	60	60	54	64
Somewhat dissimilar to the content of state test	6	7	7	6	10
Very dissimilar to the content of state test	2	4	2	3	6

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Figure 9 shows how the proportions of H/H and M/H teachers differ from that of the M/L teachers on the similarity of test preparation material to test content. The "very similar" option for both is about .4 of a standard deviation from that of M/L teachers. Both Table 30 and Figure 9 show that the higher the stakes associated with the test, the more likely teachers are to use test preparation material that is very similar to the test content.

Figure 9.
Test Preparation Content: H/H and M/H vs. M/L Stakes States

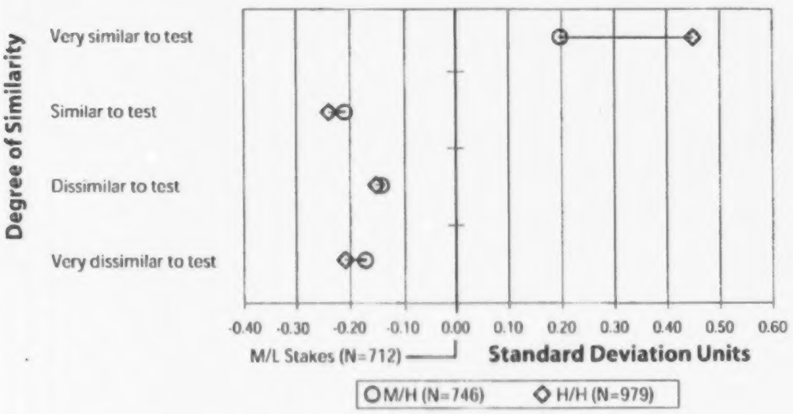


Table 31 shows the percentages for the four options in Item 65 by stakes level and grade. For the option "very similar," Table 31 shows that more elementary teachers in the H/H states (40%) and fewer in the M/L states (19%) so characterized the content of their test preparation materials. For the "similar" option, fewer elementary teachers in the H/H states (54%) and more in the M/L states (65%) so indicated. While the percentages choosing the "dissimilar" and "very dissimilar" options were quite small, fewer elementary teachers in the H/H states (5% and 1% respectively) and more in the M/L states (10% and 5% respectively) did so.

Table 31. Content of Test Preparation Material: Percent Reporting by Stakes Level and School Type[1,2]

Test Preparation Content	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Very similar to the content of state test	Elementary	40	28	34	38	19
	Middle	39	35	33	33	20
	High	41	26	27	39	22
Somewhat similar to the content of state test	Elementary	54	63	58	56	65
	Middle	51	55	62	54	61
	High	46	56	60	45	60
Somewhat dissimilar to the content of state test	Elementary	5	6	7	4	10
	Middle	8	7	3	8	11
	High	9	11	9	11	10
Very dissimilar to the content of state test	Elementary	1	3	1	2	5
	Middle	3	2	1	5	8
	High	4	8	4	4	8

1. Overall chi-square for each item is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Student Focus of Test Preparation

Table 32 shows teachers' responses to Item 66, "One test preparation strategy is to target specific groups of students. Please mark ALL that apply." Because teachers could select any of the five applicable options, a separate analysis was conducted for each option. For all five options, the differences among stakes levels were statistically significant at or beyond the .001 level.

Table 32. Groups Targeted for Test Preparation: Percent Reporting by Stakes Level[1,2]

Targeted Student Groups	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
No specific student groups	62	74	76	69	79
LEP or ESL students	8	3	4	5	3
SPED students	17	9	10	13	10
Students on the border of passing	25	13	11	20	4
Students on the border of moving to the next performance level	20	10	13	13	4

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

A smaller percentage of teachers (62%) in the H/H states than in the M/L states (79%) indicated that they did not target specific student groups. More H/H teachers (8%) than M/L teachers (3%) targeted Limited English Proficient (LEP) or English as a Second Language (ESL) students. More H/H teachers (17%) than M/L teachers (10%) targeted Special Education (SPED) students.

The same pattern holds for the targeting of students on the border of passing. In the H/H (25%), H/M (13%), H/L (11%) and M/H states (20%) more teachers selected this option than did teachers in the M/L states (4%). Finally, more H/H teachers (20%) than M/L teachers (4%) targeted students on the border of moving to the next performance level.

Figure 10 shows how the proportions of H/H and M/H teachers differ from that of M/L teachers for each of the options in Item 66. The biggest differences were for the two options dealing with students on the margin of a performance category — close to passing and close to moving to the next performance level.

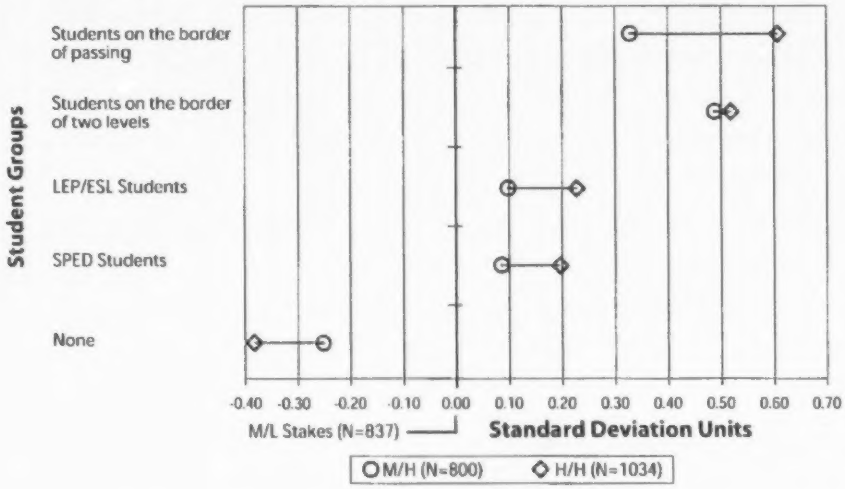


Figure 10.
Target of Test Preparation: H/H and M/H vs. M/L Stakes States

Table 32 and Figure 10 reveal that a high-stakes testing program leads to an increase in the reported incidence of targeting special groups of students for test preparation. This is as expected. When the stakes are high for students or teachers, a minority of teachers – between 8% and 25% depending on the group targeted — directed test preparation at ESL, SPED, and students on the border of passing or of the next performance level. The first three groups are students most in need of help if they are to pass. Further, moving these at-risk students from the failing to the passing category will improve school and district accountability statistics. Similarly, if students on the border of moving to the next performance level are helped to do so, accountability statistics will improve.

Table 33 shows the percentages for the five targeting options of Item 66 by grade level taught. Fewer elementary teachers in the H/H states (62%) than in the M/L states (79%) chose the option "I do not target test preparation at specific groups of students." The pattern is similar for high school teachers: 63% vs. 80%. The option "I target test preparation at LEP or ESL students," was chosen more by high school teachers in the H/H states (10%) than in the M/L states (3%). While the differences are significant, the percentages choosing this option were relatively small in both the H/H and M/L states.

Table 33. Groups Targeted for Test Preparation: Percent Reporting by Stakes Level and School Type[1,2]

Targeted Student Groups	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
No specific student groups	Elementary	62	71	76	68	79
	Middle	63	78	75	73	80
	High	63	78	78	67	80
LEP or ESL students	Elementary	8	4	6	6	3
	Middle	9	1	3	3	3
	High	10	1	2	3	3
SPED students	Elementary	14	8	9	11	9
	Middle	24	11	13	12	12
	High	18	11	11	17	9
Students on the border of passing	Elementary	25	14	12	22	5
	Middle	28	11	13	15	4
	High	22	8	5	18	3
Students on the border of moving to the next performance level	Elementary	23	11	16	17	6
	Middle	22	11	13	9	2
	High	11	5	6	6	1

1. Overall chi-square for each item is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Across all grade levels the differences were quite large between teachers in H/H states and those in M/L states for the option "I target test preparation at students on the border of passing the state-mandated test." Twenty-five percent of H/H elementary teachers chose this option, compared with 5% of M/L teachers. The same pattern holds at the middle school level (H/H 28% vs. M/L 4%) and at the high school level (H/H 22% vs. M/L 3%). Likewise, the option "I target students who are on the border of moving to the next performance level" was chosen by considerably more teachers in the H/H states than by their counterparts in the M/L states, regardless of grade level: 23% of H/H elementary teachers compared with only 6% of M/L teachers. The same pattern holds at the middle school level (H/H 22% vs. M/L 2%) and at the high school level (H/H 11% vs. M/L 1%).

Unethical Test Administration Practices

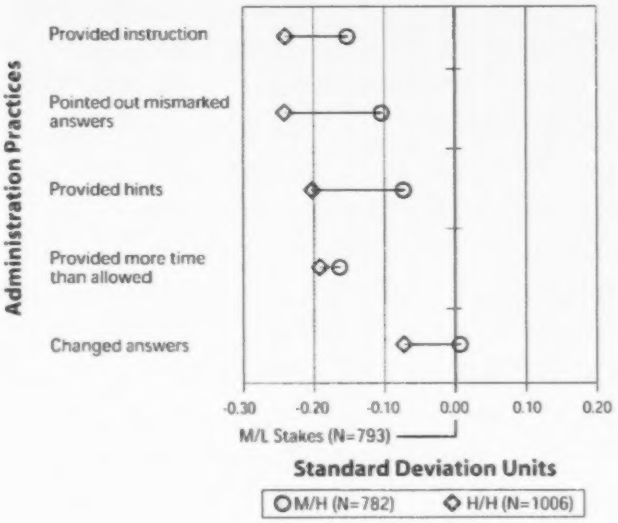
Item 67 asked, "Have you heard of any of the following activities taking place during the state-mandated test administration at your school?" Table 34 shows that while the choice of all but the option "Changed student answers on the test" was statistically significant across the five stakes levels, very few teachers selected any of the activities. Further, the activities occurred mainly in the M/L states and not in the H/H, H/M, or M/H states. (See Figure11 for a graphic display of M/L relative to the H/H and M/H teachers.)

Table 34. Teachers' Test Administration Practices: Percent Reporting by Stakes Level[1,2]

Non-Standardized Administration Practices	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Provided hints about answers	9	9	7	13	15
Pointed out mismarked items	8	10	9	12	15
Provided more time than allowed	12	15	18	13	19
Provided instruction during the test	3	7	4	5	9
Changed student answers on the test	1	2	1	2	2

1. *Overall chi-square is statistically significant (alpha = .001) only for items where any shading occurs.*
2. *Shaded values indicate significant standardized residuals (absolute values are > 3).*

Figure 11.
Unethical Test Administration Practices:
H/H and M/H vs. M/L Stakes States



A greater percentage of M/L teachers than teachers in high-stakes states indicated that they had heard of non-standardized practices taking place during test administration. Only 2% of teachers had heard that teachers had changed students' answers. The practice selected most often was giving students more time on the test, with the highest occurrence in M/L states (19%). These data can be interpreted to mean, first, that such practices are not common, and second that as the stakes increase teachers are less likely to report that they heard of the their occurrence.

Table 35 shows the percentages for the eight options by grade level in Item 67. For the option "Provided students hints about answers," fewer elementary teachers in the H/H states (11%) than in the M/L states (19%) indicated that they have heard that this occurs. Elementary teachers in the H/L states were least likely to have heard it (9%).

For the option "Pointed out mismarked items to students," fewer elementary teachers in the H/H states (9%) and more in the M/L states indicated they have heard that the practice occurs (20%). For "Provided instruction during the test," fewer elementary teachers in the H/H states (3%) and more in the M/L (9%) so reported. Again, the percentages are quite low.

Table 35. Teachers' Test Administration Practices: Percent Reporting by Stakes Level and School Type[1,2]

Non-Standardized Administration Practices	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Provided hints about answers	Elementary	11	10	9	17	19
	Middle	4	8	5	7	12
	High	7	6	6	9	8
Pointed out mismarked items	Elementary	9	12	11	14	20
	Middle	3	7	8	9	8
	High	8	6	4	7	9
Provided more time than allowed	Elementary	12	16	19	13	20
	Middle	11	14	19	13	19
	High	13	12	15	13	17
Provided instruction during the test	Elementary	3	8	4	6	9
	Middle	5	6	3	4	8
	High	4	7	4	5	11
Changed student answers on the test	Elementary	1	3	2	3	3
	Middle	1	1	1	3	0
	High	1	0	0	0	2

1. Overall chi-square for each item is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Schoolwide Motivational Practices

Item 68 presented teachers with various motivation strategies and asked, "Does your school rely on any of the following strategies to influence students to do their best work on the state-mandated test? Mark all that apply." Each of the 12 practices listed in Item 68 was examined for statistical significance across the five stakes levels. Table 36 and Figure 12 display the results of the analysis.

Table 36. Schoolwide Motivational Strategies: Percent Reporting by Stakes Level[1,2]

Motivational Strategies	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Discuss importance of good performance	72	80	82	70	66
Hold assemblies to motivate students	31	22	27	17	12
Publicly recognize students for good performance	31	28	23	22	18
Schedule special activities (e.g. pizza party, field trips)	27	28	30	20	16
Provide free time as a reward to students	14	13	19	13	11
Link performance to eligibility in extracurricular activities	7	3	3	2	2
Give prizes to reward students	19	15	16	12	6
Require/recommend summer school	43	23	23	42	8
Retain students in grade	25	11	7	11	3
Use scores for assigning grades	8	2	3	2	2
Place students in classes	34	17	20	17	15
Exempt students who do well from required course work	3	2	1	5	1

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

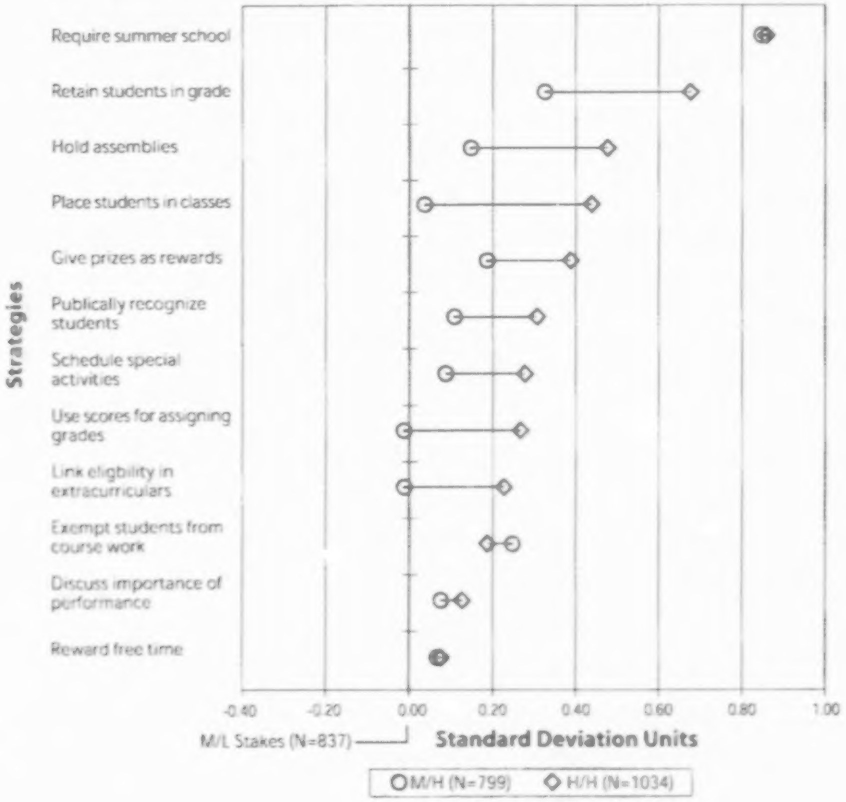
There are significant differences across the five stakes levels for all 12 strategies. Teachers in the H/H states were more likely than teachers from M/L states to choose the following:

- Discuss the importance to the school of good performance on the test (72% vs. 66%)
- Require or recommend summer school (43% vs. 8%)
- Place students in classes (e.g., honors, remedial) (34% vs. 15%)
- Publicly recognize students for good performance (31% vs. 18%)
- Hold assemblies to motivate students (31% vs. 12%)
- Schedule special activities (27% vs. 16%)
- Retain students in grade (25% vs. 3%)
- Give prizes to reward students (19% vs. 6%)

- Provide free time as a reward to students (14% vs. 11%)
- Use scores when assigning report card grades (8% vs. 2%)
- Link performance to eligibility for participation in extracurricular activities (e.g., athletics, clubs) (7% vs. 2%)
- Exempt students who do well from required course work (3% vs. 1%)

Figure 12 shows how far the responses of H/H and M/H teachers diverge from those of M/L teachers. A clear pattern emerges. Teachers from H/H states are more apt than their counterparts from M/L states to use motivational strategies to improve test performance. M/L teachers did report using some of these strategies with their students, but much less often than H/H teachers.

Figure 12.
Use of Schoolwide Motivational Strategies: H/H and M/H vs. M/L Stakes States



The increased use in high-stakes testing situations of two of these strategies is troubling: requiring summer school and retaining students in grade. In some situations, summer school could amount to little more than very intense test-specific preparation rather than skill development. Retention in grade is worrisome for at least two reasons. First, the literature is clear that grade retention increases the likelihood that a student will eventually drop out of school (see for example Clarke, Haney, & Madaus, 2000; Fassold, 1996; Heubert & Hauser, 1999; Kreitzer, Madaus, & Haney, 1989; Madaus & Greaney, 1985). Second, grade retention often results in the student receiving for another year the same sort of instruction that was unsuccessful once. For both these reasons, the motivational effect of retention is questionable. Table 37 shows the percentages for the 12 options by grade level in Item 68.

Across all grade levels, teachers in the H/H states are more likely than teachers in the M/L states to choose the following:

- Require or recommend summer school (45% of the elementary teachers, 47% of the middle school teachers, and 33% of the high school teachers in H/H states vs. 9%, 7%, and 6% in M/L states)
- Retain students in grade (26%, 31% and 18% in H/H states vs. 3%, 2% and 4% in M/L states)
- Place students in classes, e.g. honors, remedial (29%, 46% and 36% in H/H states vs. 17%, 12% and 12% in M/L states)

More elementary teachers in H/H states than in M/L states chose the following:

- Hold student assemblies to motivate students (32% of the elementary teachers in the H/H states vs. 10% of those in the M/L states)
- Publicly recognize students for good performance (28% in H/H vs. 17% in M/L)
- Schedule special activities (25% in H/H vs. 15% in M/L)
- Link performance to eligibility for participation in extracurricular activities (5% in H/H vs. 1% in M/L)
- Give prizes to reward students (17% in H/H vs. 6% in M/L)
- Use scores when assigning report card grades (7% in H/H vs. 2% in M/L elementary teachers)

Table 37. Schoolwide Motivational Strategies: Percent Reporting by Stakes Level and School Type[1,2]

Motivational Strategies	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Discuss importance of good performance	Elementary	69	79	80	67	62
	Middle	81	81	84	76	69
	High	71	82	86	72	73
Hold assemblies to motivate students	Elementary	32	23	25	18	10
	Middle	35	21	36	18	14
	High	23	20	25	16	14
Publicly recognize students for good performance	Elementary	28	30	22	23	17
	Middle	40	24	29	20	14
	High	30	26	21	22	22
Schedule special activities (e.g. pizza party, field trips)	Elementary	25	30	29	18	15
	Middle	35	25	36	22	20
	High	27	25	29	21	17
Provide free time as a reward to students	Elementary	16	17	24	15	12
	Middle	15	9	16	9	14
	High	5	7	10	11	6
Link performance to eligibility in extracurricular activities	Elementary	5	4	3	0	1
	Middle	10	4	5	3	5
	High	9	1	3	4	3
Give prizes to reward students	Elementary	17	17	16	14	6
	Middle	24	11	17	9	7
	High	19	13	12	8	7
Require/recommend summer school	Elementary	45	32	29	46	9
	Middle	47	18	22	42	7
	High	33	7	10	33	6
Retain students in grade	Elementary	26	15	6	12	3
	Middle	31	10	11	13	2
	High	18	1	3	7	4
Use scores for assigning grades	Elementary	7	2	3	3	2
	Middle	5	1	5	1	3
	High	14	2	3	1	3
Place students in classes	Elementary	29	18	21	12	17
	Middle	46	19	25	22	12
	High	36	12	12	23	12
Exempt students who do well from required course work	Elementary	1	1	0	1	0
	Middle	3	2	2	2	2
	High	10	3	3	16	2

1. Overall chi-square for each item is statistically significant (alpha = .001) only for items where any shading occurs.

2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Summary

The data on test preparation show that teachers in high-stakes testing states are more likely than are teachers from states without such programs to engage in test preparation earlier in the school year; spend more time on this activity; target special groups of students for more intense preparation; use materials that more closely resemble the test; and use more motivational tactics.

The grade-level data show that more teachers in high-stakes states than in low-stakes states, regardless of grade level, report that they the use commercially or state-developed test-specific preparation materials or released items from the state test. The number of hours given over to test preparation is higher for all teachers across grade levels in the high-stakes than the M/L situations. However, elementary teachers in high-stakes situations are more likely to report spending more time on test preparation than their secondary school counterparts. Further, elementary teachers across stakes levels were more likely to report that they engaged in test preparation throughout the year than were middle or high school teachers. Also, elementary teachers in the H/H states were twice as likely as those in M/L states to report that their test-preparation content was very similar to the content of the test. They were also four times more likely to report targeting test preparation at students on the border of passing or moving to the next performance level than their M/L counterparts.

When asked whether summer school should be required or recommended as a motivational strategy, close to 45% of elementary and middle school teachers and a third of secondary teachers in the H/H states responded affirmatively. Less than 10% of teachers across all levels in the M/L stakes states so reported. Retention in grade was selected by 1 in 4 elementary teachers, close to a third of middle school teachers, and 1 in 5 high school teachers in H/H states while the percentages in the M/L states never reached 5% across grade levels.

These data on test preparation practices need to be interpreted in light of other sections of this report before a value judgment on the appropriateness and efficacy of the various practices is made. However, experience with high-stakes tests dating back to before the 19th century indicates that there are real dangers associated with test preparation practices (see for example Greaney & Kellaghan, 1996; Madaus & Greaney, 1985; Madaus, 1988; Madaus & Kellaghan, 1992; Shepard, 2001; Smith & Rottenberg, 1991). The data from this survey illustrate the strong relationship between the stakes associated with the test and the use of various test preparation practices and are a cautionary tale showing that these historical dangers remain a real possibility.

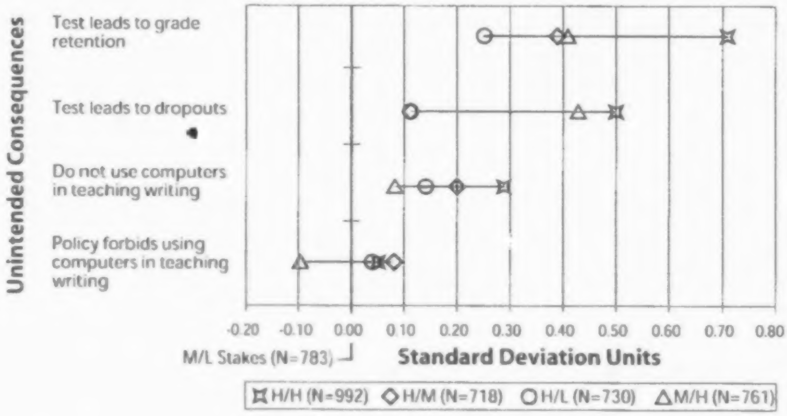
VII. Unintended Consequences of the State Test

Up to this point we have discussed teacher responses mostly in terms of directly observable effects associated with state testing programs, such as the impact of testing on curriculum or classroom assessment practices. In addition to these effects, research also suggests a potential link between the use of state tests and several unintended consequences such as high school dropouts and grade retention (see for example Clarke, Abrams, & Madaus, 2001; Clarke, Haney, & Madaus, 2000; Greaney & Kellaghan, 1996; Haney, 2000; Heubert & Hauser, 1999; Jacob, 2001; Madaus, 1988; Madaus & Kellaghan, 1992; Reardon, 1996). In order to assess whether teachers' perceptions of unintended consequences varied by the relationship between stakes for districts, schools, and teachers and the stakes for students, we asked teachers to indicate the extent of their agreement to four statements:

- State-mandated test results have led to many students being retained in grade in my district. (Item 48)
- State-mandated test results have led many students in my district to drop out of high school. (Item 46)
- Teachers in my school do NOT use computers when teaching writing because the state-mandated writing test is handwritten. (Item 6)
- My school's (district's) policy forbids the use of computers when teaching writing because it does NOT match the format of the state-mandated writing test. (Item 18)

Before we investigate responses to each item in detail, it is useful to get an overall sense of the data. Figure 13 provides a picture of response patterns across items that reflect unintended consequences of testing systems. To create this graph, the "strongly agree" and "agree" categories have been collapsed into a single "agree" category. The percentages are converted to standard deviation units, with the M/L stakes category used as a baseline. In this way, we can compare the level of agreement for each item across stakes levels. For example, if we focus on teachers' responses to the statement "State-mandated test results have led to many students being retained in grade in my district," we can see that there is a difference of .7 standard deviation units between the H/H and M/L categories. Teachers in the H/H group expressed greater agreement than their counterparts in M/L states.

Figure 13.
Agreement for Unintended Consequences: H/H, H/M, H/L, M/H vs. M/L Stakes States



The disparity in levels of agreement is largest between the H/H group and the M/L group concerning the impact of the state test on grade retention (Item 48). Given that there is a difference in perception regarding teachers' beliefs about retention, the response pattern for this item makes sense; we would expect to see increased retention rates where there is greater pressure on schools and students to do well on state-mandated test that is, in a high-stakes environment. In the graph we see that the groups with the highest level of agreement also have high stakes for students. As the stakes levels fall for students, so does the level of agreement — almost a full standard deviation when comparing the H/H and M/L groups.

When teachers were asked whether state-mandated test results led many students to drop out of high school, responses showed a much larger gap between the two groups with high stakes for students (H/H and M/H) and the rest of the stakes levels. Again, this was expected, since this item deals with the indirect impact of testing on students; that is, if testing policy contributes to dropouts, it would be more likely to do so in high-stakes environments for students.

Overall, teachers' responses across stakes levels when asked about school or district policies relating to using computers when teaching writing were consistent (Item 18). Teachers at H/H, H/M, H/L stakes levels responded similarly to M/L teachers — as suggested by the small standard deviations — that their school or district had a formal policy that forbade the use of computers to teach writing because students' responses on the state test are handwritten. However, the graph illustrates greater disparity when teachers responded to a more general question about computer use in writing instruction (Item 6).

Item-Level Results by Stakes Level and School Type

Impact on Grade Retention

Across stakes levels, teachers' responses varied with respect to the impact of the state test on grade retention. When stakes for schools and/or teachers are held constant, there is a dip in agreement with the statement as the stakes fall for students. For example, 27% of teachers in the H/H category agreed that the state test has led to grade retention, as compared with 9% in the H/L category. This is also true when schools and teachers face moderate stakes while the stakes for students vary; in this case, 14% of M/H teachers vs. 3% of teachers in M/L states indicated that the state test has influenced how many students in their district are retained in grade.

Table 38. Test Leading to Grade Retention: Percent Agreement by Stakes Level and School Type[1,2,3]

School Type	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
All teachers	27	14	9	14	3
Elementary	26	19	10	16	2
Middle	30	10	10	14	4
High	27	5	6	11	5

1. Overall chi-square for results by stakes and by stakes and grade level is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).
3. The strongly agree and agree response categories were collapsed into general-agreement responses.

When we examine the response pattern by school type across stakes levels, the highest percentage of agreement occurs when stakes are high for both schools and students. Moreover, far more teachers in H/H, H/M, H/L and M/H states than in M/L programs reported that the state test has led to grade retention in their district (27%, 14%, 9%, 14% vs. 3%, respectively). Because the issue here is retention rather than graduation, it makes sense that stakes would have a similar impact across grade levels. It is interesting to note that in the H/M testing program category, 19% of elementary teachers agreed with the statement while only 5% of high school teachers did so. Overall, most teachers across stakes levels and school types disagreed with the statement that the state test increased grade retention in their district.

Impact on Dropout Rates

The overall pattern for Item 46, which asked teachers to indicate the extent of their agreement with the statement "State-mandated testing has caused many students in my district to drop out of high school," was that a substantial majority of teachers across stakes level disagreed (H/H 72%, H/M, 87%, H/L, 87%, M/H 75%, M/L 90%). However, their responses do reveal interesting variations across stakes levels. When we collapse the response categories into "agree" and "disagree," we find that high stakes for students correspond with higher levels of agreement. For example, 28% of teachers in H/H states and 25% in M/H states agreed with the statement, as compared with 10% in M/L states. Similarly, teachers in states with high stakes for students were more likely to agree that the state test influenced students' decisions to drop out of school than teachers from states with high stakes for schools (H/M 13%, H/L 13%).

Table 39. Test Leading to Dropping Out: Percent Agreement by Stakes Level and School Type[1,2,3]

School Type	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
All teachers	28	13	13	25	10
Elementary	28	17	15	27	10
Middle	29	9	13	27	9
High	28	6	8	20	10

1. Overall chi-square for results by stakes and by stakes and grade level is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).
3. The strongly agree and agree response categories were collapsed into general-agreement responses.

Teachers' responses by school level within the various types of testing programs mirror the response pattern at the stakes level. Within school type across stakes levels, teachers' responses in states with high stakes for students are similar. Elementary, middle and high school teachers who report agreement in the largest percentages are associated with high stakes for students. As grade level increases in states with moderate or low stakes, fewer teachers reported that the state-mandated test contributes to student dropout rates. Generally across all grade levels, elementary teachers report in the largest percentages that they see the state test influencing students' decisions to drop out of high school.

The data also indicate that while most teachers disagree that dropping out has increased, disagreement runs higher in non-high-stakes environments for students. Since it is logical that high stakes lead to high pressure, this makes sense. What is particularly interesting is that the pattern is fairly consistent across school types. Presumably, elementary teachers do not contend with issues related to perseverance or graduation to the same degree as high school practitioners. Yet in the M/L stakes states, more elementary and middle school teachers indicated that state-mandated tests increased dropout rates than did high school teachers.

Impact on the Instructional Use of Computers

In addition to items directly related to grade retention and students' remaining in high school, several items addressed the impact of the use of technology in instruction. Teachers were asked to indicate the extent of their agreement with the following statements: "Teachers in my school do NOT use computers when teaching writing because the state-mandated writing test is handwritten" (Item 6) and "My school's (district's) policy forbids using computers when teaching writing because it does NOT match the format of the state-mandated writing test" (Item 18).

Teachers' responses to Item 6 indicate that the impact on computer use varies depending on the stakes level. Teachers in high-stakes states showed the greatest level of agreement; the extent of agreement decreased as stakes for students become less severe. This makes sense, since teachers in high-stakes situations might be more inclined to try to "acclimate" students to the format of the test, as suggested by previous survey results.

The responses to Item 6 indicated that a substantial majority of teachers, roughly two-thirds, at each stakes level disagreed with the statement. However, as is shown in Table 40, teachers reporting that the test format limited the classroom use of computers in teaching writing were more likely be from H/H (33%) rather than M/L states (20%).

Table 40. Computer Use Precluded by Test Format: Percent Agreement by Stakes Level and School Type[1,2,3]

School Type	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
All teachers	33	29	26	24	20
Elementary	35	34	31	25	24
Middle	34	23	26	25	15
High	25	20	14	18	13

1. Overall chi-square for results by stakes and by stakes and grade level is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).
3. The strongly agree and agree response categories were collapsed into general-agreement responses.

With regard to school type, it is clear that most teachers at each level did not agree that the format of the test influenced their colleagues' use of computers in teaching writing. However, as is also apparent from Table 40, elementary teachers were more likely to report agreement than their middle or high school counterparts, particularly in H/M, H/L, and M/L stakes states. In contrast, teachers' responses by grade level were fairly consistent in states with high stakes for students (H/H and M/H), particularly in elementary and middle schools. About 35% of these teachers in H/H and 25% in M/H states indicated that teachers at their school do not use computers to teach writing because the state test requires handwritten responses. Teachers' response patterns differed most by stakes level and in middle school grades. For example, the difference between the H/H and M/L categories is 10 percentage points larger for middle school (34% vs. 15%) than for elementary school (35% vs. 24%). Generally, as grade level increases, fewer teachers reported that computers are not used in teaching writing because of the format of the state test; and agreement decreases with decreasing stakes.

Item 18 examined formal policies at the district or school level on the use of technology relative to the format of the state test. The response patterns indicate that an overwhelming majority of teachers, roughly 95%, disagreed that a district or school policy bans computer use in teaching writing because of the format of the test (see Table 41). From this, we can infer that a formal district or school policy limiting the instructional use of computers is not a common practice.

Table 41. Policy Ban on Computer Use in Writing Instruction: Percent Agreement by Stakes Level and School Type[1,2,3]

School Type	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
All teachers	5	5	5	2	4
Elementary	5	6	5	2	5
Middle	5	5	4	3	2
High	5	3	3	3	3

1. Overall chi-square for results by stakes and by stakes and grade level is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).
3. The strongly agree and agree response categories were collapsed into general-agreement responses.

Summary

The data presented in this section indicate that most teachers disagree with the sentiments expressed in the four relevant items. The best example of this is Item 18. When asked whether school policy forbids the use of computers in teaching writing because of the format of the state test, teachers' level of disagreement was almost identical among stakes levels and school types.

However, while most teachers disagreed with the statements discussed in this section, there is some evidence that the level of agreement does vary by the stakes for students. In some cases, the difference between the H/H category and the M/L category was fairly pronounced. For instance, 34% of H/H middle school teachers vs. 15% of M/L teachers, agreed that "teachers in their school do not use computers when teaching writing because of the format of the state-mandated test" (Item 6). This disparity also appears in the items on the impact of the state test on grade retention and dropping out. It is important not to discount the perspectives of teachers who believe that state-mandated testing has an indirect impact on schools. While they may not be in the majority, their numbers are not negligible.

VIII. Use of Test Results

In this section, we discuss teachers' responses to items dealing with various uses of state-mandated test results. These uses range from making decisions about individual students to making judgments about teachers, schools, and/or districts. We asked teachers how test results were used in their district and how appropriate they found these uses. The influence of test results on teaching and the utility of test score reports are also examined, in addition to the adequacy of professional development opportunities related to the state testing program.

Teachers' Views on Accountability

State testing results have been and are being used for decisions about students, teachers, administrators, schools, and school districts. Item 61 on the survey asked teachers about the appropriateness of uses such as accountability, placement or grouping of students, and evaluation of programs (see Table 42). Since these uses are prevalent, all educators and policymakers should understand teachers' views on this topic. Areas judged inappropriate by large numbers of teachers deserve closer scrutiny; those judged appropriate have the best chance of affecting instructional practice. Teachers at different grade levels may view a certain use differently even if the stakes attached are the same, and these disparities elucidate how various uses play out in different testing programs.

Overview of Accountability

Item 61 comprises 17 sub-items representing ways in which test results are used to hold schools, teachers, and students accountable for performance on the state test. For an overview of teachers' perceptions, factor analytic techniques were used to create scales. The analysis yielded three scales, each organized around a different unit of accountability. The items composing the first scale all relate to school accountability; those in the second scale to student accountability; and those making up the third scale to teacher/administrator accountability. Table 42 presents the items that compose each scale (the technical information related to these analyses is presented in Appendix E, Table E13).

Table 42. Items Comprised by School, Student, and Teacher/Administrator Accountability Scales

Item 61: The following is a list of ways in which state-mandated test results are used for each item. Please indicate how appropriate you feel the specific use is.	School Accountability Scale	Student Accountability Scale	Teacher/Admin. Accountability Scale
Evaluate charter schools	X		
Evaluate voucher programs	X		
Hold the district accountable	X		
Hold schools accountable	X		
Award school accreditation	X		
Place schools in receivership	X		
Rank schools publicly	X		
Place students in special education		X	
Place students in gifted programs		X	
Promote/retain students in grade		X	
Remediate students		X	
Group students by ability in grade		X	
Graduate students from high school		X	
Award teachers/admin. financial bonuses			X
Reward schools financially			X
Evaluate teacher/admin. performance			X
Fire faculty/staff			X

Scale Score Results

Individual items were coded 1 for "very inappropriate" to 4 for "very appropriate"; thus a higher value represents greater appropriateness. The scores for each scale are on the same 1 to 4 metric (obtained by taking the mean of the responses for those items). Scores on each scale were examined for differences by stakes level (H/H, H/M, H/L, M/H, and M/L) and school type (elementary, middle, high school) using analysis of variance or ANOVA. Complete ANOVA tables are presented in Appendix E, Tables E14-E16.

Scores on all three scales differed significantly by stakes level, but not by type of school, suggesting that teachers' perceptions of the appropriateness of the uses of test results depend largely on the type of testing program. Table 43 presents the mean scale scores by stakes level for each scale. On average, teachers in all groups view using state test results for school accountability as "moderately inappropriate" (see Table 43). Teachers in states with high stakes for schools, teachers, and/or districts and for students differed from all the other stakes-level groups in having a higher average score: while they also viewed this use as "moderately inappropriate," they did so with a score of 1.99, whereas their counterparts were more negative (scores ranging from 1.72 to 1.84).

Further, teachers in H/L states scored higher (1.84) than teachers from M/L states (1.72); that is, they viewed the use of state-mandated test results for school accountability as less inappropriate than did the latter group. But all teachers, on average, viewed this use as inappropriate.

Table 43. Means on the Accountability Scales by Stakes Level

Stakes Level	School Accountability Scale	Student Accountability Scale	Teacher/Admin. Accountability Scale
H/H	1.99	2.52	1.55
H/M	1.81	2.25	1.33
H/L	1.84	2.22	1.41
M/H	1.75	2.28	1.27
M/L	1.72	2.24	1.29

On average, all teachers viewed the use of state test results for student accountability as being "moderately inappropriate" to "moderately appropriate," tending toward the latter. Teachers in the H/H states again differed from the other four stakes-level groups; none of those four groups differed from each other. The mean of roughly 2.52 for the H/H group places their views midway between "moderately inappropriate" and "moderately appropriate." The scores for the other four groups place them in the same range but somewhat closer to "moderately inappropriate" (see Table 43). Thus, teachers seeing the highest-stakes use of state tests were basically neutral about the appropriateness of their use for student accountability; teachers in all other groups were somewhat negative.

All teachers viewed the use of state test results for teacher/administrator accountability scale as inappropriate, on average between "moderately" and "very inappropriate" (see Table 43). Scores for teachers in H/H states once again differed from those of teachers in the other four groups. The scale score of 1.55 for teachers in the H/H group places their views midway between "moderately" and "very inappropriate" on this scale (see Table 43). Teachers in the other four groups are in the same range but tend to be more negative (ranging from 1.27 to 1.41).

On the same accountability scale, scores for teachers in the H/L group also differed from those of teachers in the M/H and M/L groups (1.41 vs. 1.27 and 1.29), falling more toward the midpoint between "very" and "moderately inappropriate," whereas teachers in the groups with moderate stakes for schools and/or districts tended more toward "very inappropriate." Again, it should be noted that all groups were, on average, in the "moderately" to "very inappropriate" range. Teachers in general viewed this use (teacher and administrator accountability) as the least appropriate of the three uses examined by this question.

In summary, all teachers, on average, were neutral regarding the use of state test results for student accountability and found their use for school accountability "moderately inappropriate," and that for teacher/administrator accountability "moderately" to "very inappropriate." Teachers in the H/H group consistently had higher scores for all three uses, viewing the accountability uses of state tests as somewhat less inappropriate than all other teachers.

On average, teachers in the H/H group felt that the use of state test results for student accountability was the most appropriate of the three (having a score between "moderately appropriate" and "moderately inappropriate," a neutral view). They found their use for teacher/administrator accountability the least appropriate (with a score between "moderately" and "very inappropriate"). The responses to Item 61 on the survey suggest that where the stakes are the highest, teachers view the use of test results for accountability at all levels somewhat more favorably (or at least less unfavorably). However, their views still fall into the neutral to unfavorable range, relative to those of teachers in states where the stakes are not as high. There are many possible reasons for this; we will put forth some of these possibilities after discussing the item-level results below.

Item-Level Results

In an effort to further understand the results presented thus far for Item 61, let us look at the responses to the individual items rated "very inappropriate" to "very appropriate" that made up the scales discussed in the previous section. When one examines teachers' responses by stakes level (H/H, H/M, H/L, M/H, and M/L) and school type (elementary, middle, and high school), a very consistent pattern emerges. Significantly more teachers in H/H states viewed the use of state test results as appropriate or very appropriate on all of the 17 items rated than did their counterparts in states with the other four configurations of stakes.

Further, a greater percentage of elementary school teachers in the H/H states viewed the use of state test results as "very appropriate" than did their counterparts at the other stakes levels. In other words, the use of state test results was rated "very appropriate" significantly more often by elementary teachers in the H/H group. Again, in some instances these percentages are small (under 10%) and differences should be interpreted with caution. Response patterns for an exemplary item from each of the three areas of accountability (school, student, teacher/administrator) discussed in the previous section may help to further clarify this point.

School Accountability

The item used to illustrate teachers' perceptions of school-level accountability is the item "hold schools accountable." Table 44 shows the range of responses by stakes level. Seven percent of teachers in the H/H group chose "very appropriate," demonstrating the point made above: significantly more teachers in the H/H category viewed using state-mandated test results to hold schools accountable as "very appropriate." The percentages in this category across all groups are small (7% or less in all cases), but the H/H group has a noticeably greater percentage than any other group. Note that the majority — roughly two-thirds of all teachers — viewed the use of test results to hold schools accountable as inappropriate.

Table 44. Use of Test Results for School Accountability: Percent Reporting by Stakes Level[1,2]

Use Considered	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Very appropriate	7	3	4	2	2
Moderately appropriate	32	31	35	30	33
Moderately inappropriate	33	33	32	37	36
Very inappropriate	28	34	30	31	30

1. Overall chi-square is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Table 45 presents teachers' responses to this item by grade level within stakes level. The data suggest that while middle and high school teachers' responses are consistent across the different types of testing programs, elementary teachers' views differ. Significantly more elementary teachers in H/H stakes states found using results to hold schools accountable to be very appropriate (even though the large majority of teachers regard it as inappropriate).

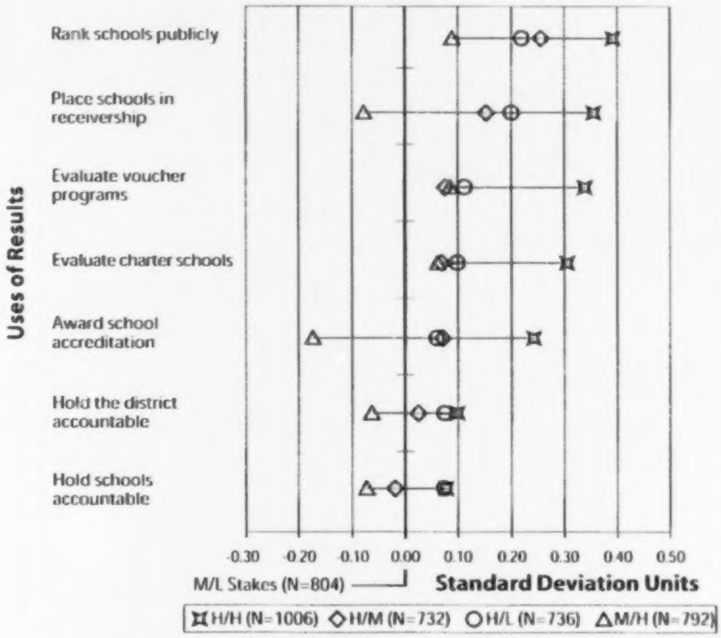
Table 45. Use of Test Results for School Accountability: Percent Reporting by Stakes Level and School Type[1,2]

Use Considered	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Very appropriate	Elementary	7	3	4	1	2
	Middle	8	4	3	2	2
	High	5	3	4	2	3
Moderately appropriate	Elementary	30	32	36	30	34
	Middle	33	31	35	26	35
	High	39	27	32	33	28
Moderately inappropriate	Elementary	33	32	30	40	35
	Middle	35	31	35	37	38
	High	31	35	32	31	36
Very inappropriate	Elementary	30	34	30	29	29
	Middle	25	33	28	35	26
	High	25	34	31	34	33

1. Overall chi-square is statistically significant (alpha = .001) only for elementary school.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Figure 14 represents all of the items for the school accountability area by showing in standard deviation units how each stakes level compares with the M/L group. One can see that the H/H group ranges from .1 to .4 standard deviation units above the baseline M/L group. The M/H group is .1 to .2 standard deviation units below the M/L for a number of the items.

Figure 14.
Appropriateness of Using Test Results for School Accountability: H/H, H/M, H/L, and M/H vs. M/L



Student Accountability

The exemplary item chosen for the student accountability area is "promote or retain students in grade." Table 46 shows teachers' responses to this item by stakes level. As in Table 44, one sees that significantly more teachers in H/H states reported that promotion or retention decisions about students based on test results are very appropriate (11%) than did teachers in states with stakes of lesser consequence (roughly 4% to 6%). All of these percentages are small and differences need to be interpreted with caution. Other cells with differences in percentages pertain to the H/H column also. These are the "moderately appropriate" and "very inappropriate" responses. The percentages are 30 and 26 respectively. Thus, for this item greater percentages of teachers in the H/H category chose the "appropriate" end of the scale (roughly 40%) more often than any of the other groups (roughly 25%), and the "very inappropriate" response (roughly 26%) less often than any of the other groups (roughly 40%). Most teachers viewed this use as inappropriate.

Table 46. Use of Test Results to Promote or Retain Students: Percent Reporting by Stakes Level[1,2]

Use Considered	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Very appropriate	11	5	4	6	5
Moderately appropriate	30	20	22	21	20
Moderately inappropriate	33	38	31	35	34
Very inappropriate	26	38	43	39	43

1. Overall chi-square is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Different patterns for teachers' responses emerged across stakes and grade levels with regard to the use of test results to make decisions about grade promotion or retention. Both elementary and high school teachers' responses differ by stakes level, while middle school teachers' views are reasonably consistent (see Table 47). Almost twice as many elementary teachers in H/H states found using test results to promote or retain students in grade to be very appropriate, as did their counterparts in other types of testing programs. Similarly, high school teachers in H/H states were also more apt to view using test results in this manner as very appropriate in comparison with those at different stakes levels. Conversely, teachers in M/H states were significantly more likely to view using test results to promote or retain students as "very inappropriate."

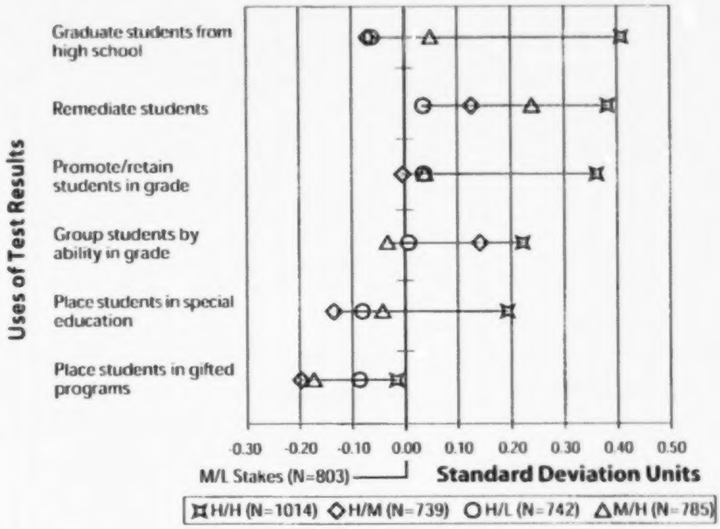
Table 47. Use of Test Results to Promote or Retain Students: Percent Reporting by Stakes Level and School Type[1,2]

Use Considered	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Very appropriate	Elementary	12	5	2	5	5
	Middle	10	4	6	7	4
	High	11	4	6	5	6
Moderately appropriate	Elementary	27	20	20	16	18
	Middle	29	20	26	22	22
	High	37	19	24	32	23
Moderately inappropriate	Elementary	33	40	31	38	34
	Middle	34	34	32	30	30
	High	29	33	30	32	34
Very inappropriate	Elementary	27	35	46	41	43
	Middle	27	42	36	42	44
	High	23	45	40	32	37

1. Overall chi-square is statistically significant (alpha = .001) only for elementary school.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Figure 15 shows for each item in the student accountability area how each stakes level group differs from the M/L group in standard deviation units. The H/H group is .2 to .4 standard deviation units above the M/L group for five of the six items.

Figure 15.
Appropriateness of Using Test Results for Student Accountability: H/H, H/M, H/L, and M/H vs. M/L



Teacher/Administrator Accountability

The item chosen as an example of teachers' responses in the teacher accountability area is "evaluate teacher or administrator performance." Table 48 shows the responses by stakes level. Once again, many more teachers in H/H states viewed the use of state test results to hold teachers and administrators accountable as appropriate (either "very" or "moderately appropriate") than did teachers in states with lower stakes (roughly 18% vs. 10%). Note, however, that the vast majority of teachers, regardless of stakes level, viewed this use as inappropriate, with most finding it "very inappropriate."

Table 48. Use of Test Results to Evaluate Teachers/ Administrators: Percent Reporting by Stakes Level[1,2]

Use Considered	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Very appropriate	4	1	2	2	1
Moderately appropriate	14	6	8	6	9
Moderately inappropriate	26	24	26	21	27
Very inappropriate	56	69	65	71	64

1. Overall chi-square is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Like other results by grade level, elementary teachers' response patterns to this item differ across stakes levels (see Table 49). Many more elementary teachers in H/H states regarded this use of test results as "very appropriate" compared with their counterparts in other types of testing programs. On the other hand, greater proportions of teachers in H/M (72%) and M/H states (72%) viewed this use as "very inappropriate."

Table 49. Use of Test Results to Evaluate Teachers/ Administrators: Percent Reporting by Stakes Level and School Type[1,2]

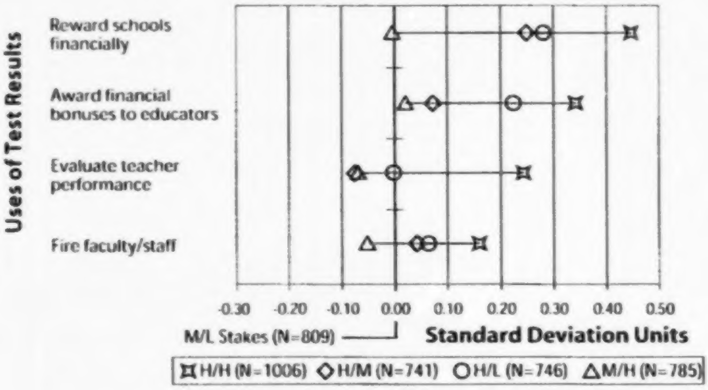
Use Considered	School Type	Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Very appropriate	Elementary	5	1	3	2	1
	Middle	1	1	1	1	1
	High	3	1	1	2	2
Moderately appropriate	Elementary	15	6	6	6	9
	Middle	13	7	5	3	7
	High	12	7	12	8	10
Moderately inappropriate	Elementary	22	21	24	21	26
	Middle	30	27	29	21	29
	High	34	27	27	23	25
Very inappropriate	Elementary	58	72	67	72	64
	Middle	56	65	64	74	64
	High	51	65	59	68	64

1. Overall chi-square is statistically significant (alpha = .001) only for elementary school.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Figure 16 shows how the stakes-level groups differ from the baseline M/L group for each item in the teacher/administrator accountability area. Again the H/H group has the largest standard deviations for all items, ranging from .2 to .45.

Figure 16.

Appropriateness of Using Test Results for Teacher Accountability: H/H, H/M, H/L, and M/H vs. M/L



Summary

More teachers in states with high-stakes — for students and schools, teachers, and/or districts — viewed the use of state test results as more appropriate for various accountability purposes that those in states with lesser stakes. Across school types, teachers view this use as quite inappropriate for teacher accountability and as moderately inappropriate for school accountability; they are essentially neutral on its use for student accountability.

The more favorable view of teachers in the H/H states could be due to their greater familiarity, and hence greater comfort, with these accountability uses. It could also be that they have simply resigned themselves to these uses and thus accept them more readily than do teachers in states where there is little or no accountability use. It is unmistakable, though, that teachers in H/H states, even though they hold a neutral to negative view about such uses of tests, generally support them to a greater extent than their colleagues in states with lower accountability levels.

District-Level Use of Test Results

As a follow-up to Item 61, which asked teachers about the appropriateness of various uses of state-mandated test results (accountability, placement or grouping of students, and evaluation of programs), Item 73 asked whether the results were in fact used to make decisions in these areas (see Table 50). In this section, we report on the prevalence of specific uses of test results.

As with Item 61, we examined the extent of use by stakes level and school type. Items in Table 50 are placed in order of reported frequency of use from highest to lowest and will be discussed in that order.

Table 50. District-Level Use of Test Results: Percent Reporting by Stakes Level[1,2]

Motivational Strategies	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Rank schools publicly	66	44	54	53	39
Hold schools accountable	63	45	57	46	35
Hold district accountable	49	39	48	40	28
Remediate students	57	30	26	45	16
Evaluate teacher/administrator	40	23	31	18	18
Place students in honors classes	33	20	26	20	23
Graduate students from high school	41	9	5	40	6
Promote or retain students in grade	30	11	5	13	5
Reward schools financially	27	13	16	6	2
Place school in receivership	16	15	18	9	1
Place in student in special education	15	9	9	7	12
Award school accreditation	18	9	19	3	4
Group students by ability	16	9	6	8	7
Award teachers/administrators financially	19	3	10	4	1
Evaluate charter schools	6	3	4	3	1
Fire faculty/staff	6	2	5	2	1
Evaluate voucher programs	2	1	2	1	1
None of the above	4	14	9	10	19

1. Overall chi-square is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Item-Level Results by Stakes Level

Roughly half of all teachers indicated that in their district state-mandated test results were used to "rank schools publicly" and "hold schools accountable." Far more teachers in H/H states (roughly 65%) said that tests were so used than did teachers in M/L states (roughly 35%). Next in frequency of use was "hold the district accountable"; 40% of all teachers reported that their district used the test for this purpose. Again, teachers in H/H states cited this use more frequently (49%) than did teachers in M/L states (28%). Teachers in H/L and H/H states reported similarly.

"Remediate students" was the next most frequent use cited by 36% of all teachers. Again, this use was most cited by teachers in H/H states (57%) and least by teachers in M/L states (16%). Roughly one-fourth of all teachers indicated that test results were used to "place students in gifted and talented/honors programs." This response was fairly uniform across stakes levels, but cited more frequently in the H/H states (33%) than at other stakes levels. Roughly one-fifth of all teachers said the tests were used to determine whether students graduate from high school. Not surprisingly, there was tremendous disparity on this use by stakes level. Roughly 40% of teachers in the H/H and M/H states, which have high stakes for students, reported this use, whereas less than 10% of those at any of the other stakes levels did so.

"Promote or retain students in grade" and "reward schools financially" were the next most frequently cited uses across all teachers (13%). Teachers in H/H states cited using test results to "promote or retain" more frequently (30%) than other teachers, especially those in M/L states (less than 5%). Interestingly, teachers in the M/H group, the other group with high stakes for students, did not cite this use with anywhere near the frequency of the H/H group (13% vs. 30%). Similarly, the H/M and H/L groups, the other two groups with high stakes for teachers, schools, and/or districts, did not report that their district used the tests to "reward schools financially" to the same extent as teachers in the H/H group did (27% vs. roughly 15%).

"Place public schools in receivership" was next on the list (12% of all teachers indicating that the test was used for this purpose). As would be expected, only 1% of teachers in the M/L group cited this use; 15% or more of the teachers in the groups having high stakes for schools, teachers, and/or districts did so. Slightly more than 10% of teachers indicated that test results were used in their district to "place students in special education" or "award school accreditation." The greatest percentage of teachers indicating that test results were used to place students in special education was the H/H group (15%); interestingly, the lowest percentage was the M/H group (7%). There was wide disparity among the groups on the use "award school accreditation," with 18% of the teachers in the H/H and H/L groups citing this use, as opposed to 3% of teachers in the M/H and M/L groups.

The remaining uses were cited by less than 10% of the teachers overall. Only teachers in the H/H group cited the next two uses, "group students by ability in grade" and "award teachers or administrators financial bonuses," with any frequency (16% and 19% respectively). Ten percent or less of all other teachers (typically only 1% to 10% for the use "award teachers or administrators financial bonuses") indicated that their districts used the tests for these purposes.

Overall, less than 5% of teachers cited each of the remaining three uses, "evaluate charter schools," "fire faculty/staff," and "evaluate voucher programs." Six percent of H/H and only 1% of M/L teachers indicated that their district used the tests to "evaluate charter schools" or "fire faculty/staff." Tests clearly are not used with any frequency (around 1% of the time) to evaluate voucher programs, perhaps because these programs are relatively rare.

In summary, we see a wide range in the reported frequency of use of state-mandated test results for the various purposes listed in this item. Teachers cite school and district accountability uses most frequently (roughly 50% of teachers report these uses). They cite evaluation of specific programs (charter schools and voucher programs) and firing faculty or staff as the least frequent uses (less than 5% of teachers report these uses). In virtually all instances, more teachers in the H/H group reported these uses of test results than did M/L teachers. For uses that pertain to students, teachers in the M/H group, the only group besides the H/H with high stakes for students, reported the use of test results with about the same frequency as H/H teachers. Uses relating to schools, teachers, and/or districts, were reported by teachers in the H/M and H/L states generally with a frequency closer to the H/H group than the M/H and M/L groups.

Item-Level Results by School Type

We also examined Item 73 by school type. The results of this analysis are presented in Table 51 and are ranked from high to low by the percentage of teachers responding that test results were used in that manner. Over half of all teachers indicated that test results are used to "rank schools publicly" (59%) and "hold schools accountable" (57%). In both instances, the lowest percentage of teachers choosing these uses was in high school.

Slightly less than half of all teachers indicated that test results are used to "remediate students" (47%) and "hold the district accountable" (45%). Again, the lowest percentage of teachers choosing these uses was among high school teachers. Roughly one-third of all teachers said test results were used in their district to "evaluate teacher or administrator performance" (33%) and "graduate students from high school" (32%). More elementary teachers (37%) and fewer high school teachers (24%) cited the former use. Not surprisingly, the reverse was true for the latter: 28% of elementary teachers and 38% of high school teachers reported that test results were used to decide on high school graduation.

About 20% of all teachers indicated that the tests were used in their district to "promote students or retain them in grade" (22%) and to "reward schools financially" (21%). As with many of the previous items, high school teachers chose these uses less often (13%) than elementary or middle school teachers.

About 12% to 15% of teachers chose the next group of uses. These were "place public schools in receivership" (15%), "award school accreditation" (14%), "award teachers or administrators financial bonuses" (13%), "place students in special education" (13%), and "group students by ability in grade" (13%). There was no difference across the three school types for placing schools in receivership. Fewer high school teachers chose each of the remaining uses. More middle school teachers chose the last three uses (awarding financial bonuses, placement in special education, and grouping students by ability).

The remaining three uses were chosen by 5% or less of all teachers: "evaluate charter schools" (5%), "fire faculty/staff" (5%), and "evaluate voucher programs" (2%). There were no differences across the three school types on these uses.

Table 51. District-Level Use of Test Results: Percent Reporting by School Type[1,2]

Use of Test Results	School Type		
	Elementary	Middle	High
Rank schools publicly	62	58	52
Hold schools accountable	60	56	48
Remediate to students	48	51	41
Hold district accountable	49	44	38
Evaluate teacher/administrator	37	32	24
Graduate students from high school	28	36	38
Place students in honors classes	33	35	12
Promote or retain students in grade	24	25	13
Reward schools financially	22	25	13
Place school in receivership	15	17	14
Award school accreditation	16	15	9
Award teachers/administrators financially	13	17	8
Place in student in special education	14	17	6
Group students by ability	13	18	7
Evaluate charter schools	5	5	4
Fire faculty/staff	4	5	5
Evaluate voucher programs	3	1	1
None of the above	6	7	11

1. Overall chi-square is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Summary

Of the various uses of statewide test results, teachers most often cited those related to holding schools and districts accountable, ranking schools, and remediating students. Fewer high school than elementary school teachers indicated that test results were used for these purposes. The only use that high school teachers chose more often than teachers at the other levels was, not surprisingly, that of basing graduation from high school on test results. Most uses were cited by less than 30% of all teachers and many by less than 15%. This pattern may be due to less awareness at the high school level than in elementary or middle schools about how test results are used in the district; or perhaps these uses are more specific to the lower grades. The latter explanation may make sense for some uses (e.g. "placement in special education" or "group by ability"), where the decisions have been made before high school and are simply carried forward independently of the state test. It makes less sense for other district uses (e.g. "rank schools publicly" or "hold schools accountable"), which should be the same across all three school types.

Influence of School and Student Results on Teaching

Teachers responded to two items about the influence of test results on their teaching, Items 69 and 70. Item 69 asks: "How often do your school's results on the state-mandated test influence your own teaching?" Item 70 asks: "How often do your students' results on the state-mandated test influence your own teaching?" A central issue surrounding state-mandated testing programs is their impact on what goes on in classrooms. The intent of these programs is, at least in part, to change instructional practices so that the state standards are taught and achieved by students. These items attempt to determine how much the results of state tests influence teaching.

The results for these two items by stakes levels (H/H, H/M, H/L, M/H, and M/L) are presented in Tables 52 and 53. Table 52 shows that for Item 69 the greatest differences are between the H/H and M/L groups. Many more H/H teachers indicate that their school's results influenced their teaching on a daily basis (40%) than did M/L teachers (10%). Conversely, a greater percentage of teachers in the M/L group reported that the school's results influenced their teaching a few times a year (24%) than did teachers in the H/H group (12%).

Table 52. Influence of School's Test Results on Teaching: Percent Reporting by Stakes Level[1,2]

Frequency	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Daily	40	23	32	27	10
Few times/week	15	12	13	11	12
Few times/month	11	15	14	13	13
Few times/year	12	21	14	21	24
Never	8	11	10	9	15
Did not receive results in time to use them	8	11	11	11	14
No results for my grade/subject	6	8	7	7	10
I should but didn't get results	0	0	1	1	2

1. Overall chi-square is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

A similar pattern emerges for Item 70 (see Table 53). A much higher percentage of teachers in the H/H group reported that their students' test results influenced their teaching on a daily basis (38%) than did teachers in the M/L stakes states (12%). The H/H group had the smallest percentage of teachers who indicated that the test results influenced their teaching a few times a year (8%), as compared with 16% of M/L teachers who so indicated. The largest percentage who reported that the results never influenced their teaching is in the M/L group (14%).

Table 53. Influence of Students' Test Results on Teaching: Percent Reporting by Stakes Level[1,2]

Frequency	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Daily	38	21	27	25	12
Few times/week	14	10	15	10	12
Few times/month	9	12	10	12	14
Few times/year	8	14	12	16	16
Never	8	12	10	8	14
Did not receive results in time to use them	14	19	17	18	19
No results for my grade/subject	10	11	8	9	11
I should but didn't get results	1	1	1	1	2

1. Overall chi-square is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Tables 54 and 55 present results for the same two items broken down by school type. For both items, there are large differences between the elementary and high school teachers. The schools' results influenced 40% of elementary teachers on a daily basis as compared with 17% of high school teachers. The students' results influenced 37% of elementary teachers on a daily basis compared with 17% of high school teachers. In both instances, middle school teachers' responses fell between the other two groups but are more similar to the responses of elementary teachers.

Table 54. Influence of School's Test Results on Teaching: Percent Reporting by School Type[1,2]

Frequency	School Type		
	Elementary	Middle	High
Daily	40	34	17
Few times/week	16	12	11
Few times/month	11	13	13
Few times/year	13	14	22
Never	6	8	18
Did not receive results in time to use them	9	10	8
No results for my grade/subject	5	7	10
I should but didn't get results	0	2	1

1. Overall chi-square is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

The flip side of these results can be seen in Table 54; 18% of high school teachers said that the school's results never influence their teaching, compared with only 6% of elementary teachers. A few other differences can be seen in Table 54. A significantly greater percentage of high school teachers reported that they do not receive the school's test results for the grade or subject they teach (10%) than did elementary teachers (5%). Although small in percentage terms, significantly more middle school teachers reported that they teach a grade or subject for which they should, but did not, receive the school's results (2%).

Returning to Table 55, we can see that a greater percentage of high school teachers reported that their student's test results influenced their teaching a few times a year (17%) or never (18%) than did elementary teachers (8% and 6% respectively). Again, middle school teachers fell between the other two groups.

Table 55. Influence of Students' Test Results on Teaching: Percent Reporting by School Type[1,2]

Frequency	School Type		
	Elementary	Middle	High
Daily	37	33	17
Few times/week	14	11	12
Few times/month	10	10	11
Few times/year	8	12	17
Never	6	8	18
Did not receive results in time to use them	17	15	13
No results for my grade/subject	10	10	10
I should but didn't get results	0	2	2

1. Overall chi-square is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Summary

Similar patterns of stakes-level differences emerge for Items 69 and 70. A much higher percentage of teachers in the H/H group reported that their students' test results influenced their teaching on a daily basis (roughly 40%) than did teachers in the M/L group (roughly 10%). The smallest percentage of teachers who said the test results influenced their teaching a few times a year was the H/H group (roughly 10%), and the largest percentage was M/L teachers (roughly 15%). Thus, the two extreme stakes-level groups (H/H and M/L) show that the results of the state test have far more influence on teaching in high- than in low-stakes states.

These two items also clearly show that state-mandated test results influence elementary teachers' instruction much more often than that of secondary teachers. This may occur because the tests now focus elementary instruction on the standards tested, giving elementary teachers, who teach a variety of subjects, much greater direction on what should be taught. These findings may also indicate that the elementary curriculum is being narrowed or shaped the most by state-mandated tests. Conversely, high school teachers' instruction may be least influenced by the state tests because these teachers have always taught a specific subject area and the test measures, for the most part, the content they were already teaching before state testing. Middle school teachers fall somewhere between elementary and high school teachers in terms of subject matter specialization, and therefore the influence of the state tests results on their instruction falls somewhere between the other two groups, although generally closer to the elementary teachers.

Classroom-Level Use of Test Results

We were interested not only in the frequency with which the school's and students' results influence classroom instruction, but also in the specific ways in which teachers use the results of the state test. Item 72 listed a series of activities and asked: "Do you use the results of the state-mandated test for any of the following activities?" (see Table 56). Table 56 presents the results for this item by stakes level; the activities are listed in order from most to least reported.

Table 56. Classroom Use of Test Results: Percent Reporting by Stakes Level[1,2]

Activities	Stakes Level				
	H/H	H/M	H/L	M/H	M/L
Plan my instruction	61	47	59	53	42
Plan curriculum	41	42	49	42	41
Select instruction materials	48	36	43	39	28
Assess teaching effectiveness	39	35	45	30	40
Give feedback to parents	41	29	37	35	32
Give feedback to students	39	24	30	27	22
Evaluate student progress	28	19	18	16	19
I didn't get results	17	17	21	21	21
Group within my class	15	5	6	8	4
Determine student grades	6	1	2	1	2
None of above	12	19	14	18	18

1. Overall chi-square is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

As can be seen from Table 56, over half (53%) of all teachers use the test results to plan their instruction. Teachers in the H/H group use them for this purpose the most (61%) and those in the M/L group the least (42%). The next two activities in terms of frequency of use are "plan curriculum" (43%) and "select instructional materials" (39%). There were no differences among stakes levels with regard to planning curriculum. Teachers in the H/H group used the results to select instructional materials the most (48 %) and teachers in the M/L group the least (28%). Thus, the top three uses of test results across all teachers are for instructional purposes.

The next most frequently reported use is "assess my teaching effectiveness" (38%). Teachers in the H/L group made use of the tests results for this purpose the most (45%) and teachers in the M/H group the least (30%). It is unclear why these two groups differ on this use.

The next two most frequently cited uses are "give feedback to parents" (35%) and "give feedback to students" (29%). The stakes-level groups did not differ from each other on giving feedback to parents, but did on giving feedback to students. Teachers in the H/H group used the results to give feedback to students the most (39%) and teachers in the M/L used them the least (22%).

The next most frequently cited use was "evaluate student progress" (20%). Teachers in the H/H group used the results the most for this purpose (28%). Next in terms of frequency was "do not get the results back in time to use them" (19%). The stakes levels did not differ in this area.

Less than 10% of teachers indicate that they used the results to "group students within my class" (8%) or "determine student grades" (in whole or in part) (3%). Teachers in the H/H group cited both of these uses (15% and 6% respectively). Teachers in the M/L group indicated using the results to group students within their class the least (4%). Decisions about individual student's placement or grades are clearly beyond the scope of what most teachers see as appropriate uses of state-mandated test results, whereas decisions about global planning of instruction are viewed as appropriate.

Results by School Type

Table 57 presents the results for Item 72 by school type. Elementary and high school teachers differed on many of the activities, with more elementary teachers indicating that they use the results than secondary teachers. Almost 62% of elementary teachers reported using the results to plan their instruction, as compared with 45% of secondary teachers. Forty-eight percent of elementary teachers reported using the results to select instructional materials; only 35% of secondary teachers did so.

Roughly 45% of elementary teachers indicated that they used the results to assess their teaching effectiveness and to give feedback to parents; the corresponding percentages for secondary teachers are 27% and 19%. Elementary teachers reported using results to evaluate student progress more frequently (28%) than did secondary teachers (15%). They also used them to group students within their class (16%) more than either middle school (8%) or secondary teachers (3%). The only use cited by a greater percentage of secondary teachers was "determining students' grades" (8%). This difference may be due to the fact that virtually all secondary teachers must give grades to their students, whereas many elementary and middle school teachers may report student achievement in other ways.

Table 57. Classroom Use of Test Results: Percent Reporting by School Type[1,2]

Activities	School Type		
	Elementary	Middle	High
Plan my instruction	62	55	45
Select instruct materials	48	41	35
Plan curriculum	41	46	38
Assess teaching effectiveness	44	32	27
Give feedback to parents	46	32	19
Give feedback to students	35	35	31
Evaluate student progress	28	24	15
I didn't get results	19	20	15
Group within my class	16	8	3
Determine student grades	3	2	8
None of the above	11	13	25

1. Overall chi-square is statistically significant (alpha = .001) only for items where any shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Summary

Teachers in the H/H group tend to use state-mandated test results to the greatest extent; those in the M/L group tend to use them the least. The greatest percentage of all teachers used the results to plan instruction or curriculum or to select instructional materials. H/H teachers used the results the most of any group to plan instruction and select materials; M/L teachers used them the least. The test results are used by about a third of all teachers to assess their teaching effectiveness and give feedback to parents or students. More teachers in the H/H group used them to give feedback to students that did teachers in other stakes-level groups; the M/L group used them the least for this purpose. More teachers in the H/H group also used the results to evaluate student progress (28%); to group students within their class (15%); and to determine students' grades (6%). It should be noted that the latter two uses are cited by a small percentage of teachers.

Clearly, the stakes attached to the state-mandated tests affect the extent to which many teachers use results for various instructional and feedback activities. When the stakes are high for students and teachers, schools, or districts, classroom teachers tend to use the results most frequently. When the stakes are low for students and moderate for teachers, schools, or districts, fewer teachers tend to use the results. For virtually all activities, less than half of the teachers indicated that they use the results, the lone exception being to plan instruction (53%). Thus, although there are differences in the degree to which teachers at different stakes levels use test results, the majority do not report using them for 7 of the 8 activities listed.

Further, very small percentages (less than 10% overall) use the results for student-specific decisions (e.g. grouping students within the class or determining student grades).

More elementary than high school teachers use the results of the state-mandated test to aid in decisions about instruction, assess their own teaching effectiveness, provide feedback to parents, evaluate students, and groups students in their class. Since elementary teachers spend most of their day with the same class, they probably get to know their students better than either middle or high school teachers, who spend far less time with their students. One might hypothesize, therefore, that elementary teachers would make less use of external test information. The survey's findings show the opposite. One possible explanation for this is that the state-mandated tests and the standards on which they are based differ the most from what elementary teachers had been doing before the testing program, so that the potential spur to change or to rely on the test is greatest for these teachers.

In general, high school teachers reported using state-mandated test results the least. These teachers generally are subject-matter-specific in their teaching and usually have a college major in their academic area. They may feel the most comfortable with the content and how they were teaching it before standards and the testing program were introduced, and therefore see less need to change their practice. Since virtually all high school teachers must grade their students, some small percentage (but a greater percentage than either elementary or middle school teachers) use the results in determining student grades.

Reporting of Test Results

The reporting of test results is often an overlooked area. Test reports that are not easily understood or that do not provide results that are useful to the intended audiences will not have much impact. For this reason, in Item 71 we asked teachers about three types of reports on student test performance: (1) the individual student's report, (2) the school report, and (3) the district report. Teachers were asked about the extent to which they agreed that each report was (1) easy to interpret and (2) provided useful information. They used a four-point Likert scale ("strongly agree" to "strongly disagree") to respond; a fifth option was ("have never seen the report"). For the purpose of the analyses here, we collapsed the two agree options ("strongly agree" and "agree") and the two disagree options ("strongly disagree" and "disagree").

Between 50% and 70% of all teachers, regardless of stakes, agreed that all the reports were easy to interpret and provided useful information (see tables 58-60). The smallest percentage of teachers (50%) agreed with these statements about the district report. This was due to the 12% to 13% of teachers who indicated that they had never seen that report, as compared with 7% to 8% of teachers who indicated that they had not seen the student or school reports.

There were very few differences among the stakes-level groups with respect to these statements. As can be seen from Table 58, more teachers in the H/L group (36%) disagreed that the individual student reports were easy to interpret than did teachers in the other groups (27% to 30%). The only difference among stakes levels for the school and district reports was that more of the M/L group indicated that they had never seen these reports than did teachers in the groups with high stakes for teachers, schools, and/or districts. Eleven percent of the M/L teachers said they never saw the school reports, as compared with 5% to 9% for the H/H, H/M, and H/L teachers (see Table 59). Roughly 17% of the M/L teachers said they had never seen the district reports; this figure was about 11% for teachers in the three groups with high stakes for teachers, schools, and/or districts (see Table 60). This finding makes perfectly good sense; where the stakes are the least severe, more teachers indicate that they have not seen the reports. Overall, however, most teachers (generally about 90%) reported having seen all the relevant reports.

Table 58. Characteristics of the *Individual Student* Reports: Percent Reporting by Stakes Level[1,2]

Student Reports		Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Are easy to interpret	Agree	67	63	59	59	62
	Disagree	27	27	36	30	28
	Never seen	6	10	6	11	11
Provide useful information	Agree	68	60	61	56	62
	Disagree	26	30	34	33	28
	Never seen	6	10	5	11	10

1. *Overall chi-square is statistically significant at (alpha = .001) only for items where any shading occurs.*
2. *Shaded values indicate significant standardized residuals (absolute values are > 3).*

Table 59. Characteristics of the *School* Reports: Percent Reporting by Stakes Level[1,2]

School Reports		Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Are easy to interpret	Agree	58	61	54	55	55
	Disagree	35	32	41	35	34
	Never seen	6	7	5	9	11
Provide useful information	Agree	67	61	63	56	59
	Disagree	27	34	32	34	30
	Never seen	6	5	5	10	11

1. *Overall chi-square for each item is statistically significant (alpha = .001).*
2. *Shaded values indicate significant standardized residuals (absolute values are > 3).*

Table 60. Characteristics of the *District* Reports: Percent Reporting by Stakes Level[1,2]

District Reports		Stakes Level				
		H/H	H/M	H/L	M/H	M/L
Are easy to interpret	Agree	50	55	49	50	51
	Disagree	39	35	40	36	32
	Never seen	11	11	11	14	17
Provide useful information	Agree	55	51	53	46	48
	Disagree	35	39	36	40	36
	Never seen	10	10	11	14	16

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Results by School Type

When Item 71 is examined by school type, we find some interesting differences. More high school teachers indicated that they have never seen the reports (see Tables 61-63): 13% said that they have never seen the student reports, 11% the school reports, and 17% the district reports. By comparison, 5% to 7% of elementary and middle school teachers said that they had never seen the student or school reports, and 9% to 13% had never seen the district reports. For three of the six statements, significantly fewer elementary teachers indicated that they had never seen the reports. The smallest percentage agreeing that the reports provide useful information are high school teachers (55% vs. 67% of elementary and middle school teachers for the student reports; 54% vs. roughly 65% for the school reports; and 46% vs. 55% for the district reports). Thus, fewer high school teachers indicated having seen the reports or finding them useful than did either elementary or middle school teachers.

Table 61. Characteristics of the *Individual Student* Reports: Percent Reporting by School Type[1,2]

Student Reports		School Type		
		Elementary	Middle	High
Are easy to interpret	Agree	67	62	60
	Disagree	28	31	27
	Never seen	5	7	13
Provide useful information	Agree	67	67	55
	Disagree	27	27	33
	Never seen	5	6	13

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3)

Table 62. Characteristics of the *School* Reports: Percent Reporting by School Type[1,2]

School Reports		School Type		
		Elementary	Middle	High
Are easy to interpret	Agree	59	55	57
	Disagree	35	38	32
	Never seen	5	7	11
Provide useful information	Agree	68	64	54
	Disagree	27	29	35
	Never seen	5	6	11

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3)

Table 63. Characteristics of the *District* Reports: Percent Reporting by School Type[1,2]

District Reports		School Type		
		Elementary	Middle	High
Are easy to interpret	Agree	51	50	49
	Disagree	40	38	33
	Never seen	9	13	17
Provide useful information	Agree	55	55	46
	Disagree	37	33	37
	Never seen	9	12	17

1. Overall chi-square is statistically significant (alpha = .001) only for item where shading occurs.
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

Summary

Fifty to 62% of all teachers either agreed or strongly agreed that both the individual student reports and the school and district reports are easy to interpret and provide useful information. The degree of familiarity with the various reports is related to the stakes attached to the results. Greater percentages of teachers in the M/L group were unfamiliar with the school and district reports than were teachers in the three groups with high stakes for teachers, schools, and/or districts (H/H, H/M, and H/L).

BEST COPY AVAILABLE

High school teachers are the least familiar with the reports; 11% to 17% indicated that they have never seen them. A smaller percentage of high school teachers agreed that the reports provide useful information. Elementary teachers are the most familiar with the reports: only 5% to 9% indicated that they had never seen them. About 13% of high school teachers said that they had never seen the individual student reports. This is a fairly large percentage. The tests are intended to provide teachers with information about their students; they cannot do this if teachers do not see the results. It could be that by the time the results come back to teachers, the tested students have moved on to the next grade. Why this would occur more often at the high school level is unclear. What is clear is that student results need to reach teachers if they are to have any impact on instruction.

Impact on Professional Development

Professional development is another area that has been influenced by the implementation of state testing programs. In an effort to gain insight into the professional resources available to teachers we asked two questions (Item 74 and 75). Item 74 asked: "Is there at least one person at your school that teachers can turn to for accurate information about the state-mandated testing program?"The vast majority of all teachers (80%) indicated that they do have someone to turn to. By stakes levels, 84% in H/H, 79% in H/M, 79% in H/L, 82% in M/H and 73% in M/L so reported. The sole difference among stakes levels occurs between teachers in the H/H group and those in the M/L group (84% vs. 73%). The other three groups fall between these extremes but closer to the H/H group. There was no difference among teachers by school type: 81% of elementary, 84% of middle, and 82% of high school teachers indicated that they had a resource person in their school. Thus, most teachers have someone knowledgeable about the testing program available to them. Where the stakes are highest, more teachers have such a person; where they are lowest, fewer have such a resource.

The second item related to professional development (Item 75) asked teachers: "How adequate has professional development in the following areas been in preparing teachers in your district to implement the state-mandated testing program?"Teachers responded by selecting one of four options ranging from "very adequate" to "very inadequate."The response option "no professional development" was also provided.

As can be seen from Table 64, most teachers viewed professional development related to the testing program to be either "adequate" or "very adequate," and professional development related to the knowledge of curriculum standards or frameworks to be "adequate" (77%). Similar numbers indicated that in-services or training on aligning classroom practices with the state test (64%) and the state standards (71%) was adequate ("adequate" or "very adequate" combined). With regard to the administration of the state test, 76% reported that the professional development was "adequate." A smaller yet sizable proportion (63%) indicated that professional development in the area of interpreting test results was also "adequate." Less than 10% of teachers indicated that there was no professional development in any of these areas, leaving 20% to 40% who feel that the professional development was inadequate ("inadequate" or "very inadequate" combined).

Table 64. Adequacy of Professional Development: Percent of Teachers Reporting

Professional Development Area	Very Adequate	Adequate	Inadequate	Very Inadequate	No Professional Development
Knowledge of state standards	21	56	15	6	2
Alignment with state standards	18	53	19	7	3
Alignment with state test	13	51	24	8	5
Test preparation strategies	10	50	27	7	7
Administration of state test	16	60	14	4	6
Interpretation of test results	9	54	23	7	8
Use of test results	6	44	32	9	9

Results by Stakes Level

The only differences among stakes levels were between teachers in the H/H group and those in the M/L group, the two extremes. Table 65 shows the professional development activities and the response categories within those activities where differences between these two groups were found. Fewer teachers in the M/L group (16%) felt that the professional development related to knowledge of the state curriculum standards or frameworks was "very adequate" than did those in the H/H group (26%). Similarly, fewer M/L teachers (8%) thought that professional development related to the alignment of classroom curriculum with the state-mandated test was "very adequate" than did H/H teachers (17%).

Table 65. Adequacy of Professional Development: Percent of H/H and M/L Teachers Reporting[1,2]

Professional Development Area	H/H	M/L
Knowledge of curriculum standards: Very adequate	26	16
Alignment of classroom curriculum to the state test: Very adequate	17	8
Test preparation strategies: Adequate or very adequate	66	43
Test preparation strategies: No professional development	4	12
Administration of state test: Very adequate	24	9
Administration of state test: No professional development	3	12
Interpretation of test results: Very adequate	14	5
Interpretation of test results: No professional development	4	14
Use of test results: Adequate or very adequate	61	42
Use of test results: No professional development	5	14

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

With respect to test preparation strategies, a significantly greater percentage of H/H teachers judged the professional development to be at least "adequate" (66%) as compared with 43% of the M/L teachers. Conversely, a greater percentage of M/L teachers (12%) said there was "no professional development" for this area than did H/H teachers (4%). A similar pattern appeared regarding professional development on the use of test results. Sixty-one percent of H/H teachers felt that professional development was at least "adequate" in this area, compared with 42% of M/L teachers. Fourteen percent of M/L teachers said there was "no professional development" in this area as opposed to 5% of H/H teachers.

When asked about professional development related to administration of the state-mandated test, significantly more H/H teachers said it was "very adequate" (24%) than did M/L teachers (9%); while 12% of M/L teachers (12%) said that there was "no professional development" in this area compared with 3% of H/H teachers. The pattern was similar with respect to professional development related to interpretation of the test results, 14% of H/H teachers judging it to be "very adequate" vs. 5% of M/L teachers. Fourteen percent of M/L teachers said there was "no professional development" in this area as compared with 4% of H/H teachers.

Thus, greater percentages of teachers in the high-stakes category consistently viewed the professional development to be adequate or very adequate than in the lower-stakes categories. Conversely, greater percentages of teachers in the lower-stakes group indicated that they received no professional development for a number of than did teachers in the higher-stakes group. The amount and adequacy of professional development appears to increase when the stakes are high for districts, schools, teachers, and students.

Results by School Type

Table 66 shows the professional development activities and the response categories within those activities where differences among school types were found. As can be seen from Table 66, significant differences occur at the high school level. The three areas are (1) test preparation strategies, (2) interpretation of the test results, and (3) use of test results. For each area, the smallest percentage judging professional development as "very adequate" was among high school teachers (4% to 9%); this contrasts with 9% to 14% of elementary and middle school teachers. These differences of roughly 5% are small in one sense, but large in a relative sense; i.e., a difference (change) of 4% to 9% more than doubles the percentage. It should be noted that the differences here are in the "very adequate" response category, that is, in the intensity of teachers' view of the adequacy of the professional development activities.

Table 66. Adequacy of Professional Development: Percent Reporting by School Type[1,2]

Professional Development Area	School Type		
	Elementary	Middle	High
Test preparation strategies: Very adequate	14	12	9
Test preparation strategies: Inadequate	23	24	30
Interpretation of test results: Very adequate	13	11	8
Interpretation of test results: No professional development	4	5	9
Use of test results: Very adequate	9	9	4
Use of test results: No professional development	6	5	10

1. Overall chi-square for each item is statistically significant (alpha = .001).
2. Shaded values indicate significant standardized residuals (absolute values are > 3).

A greater percentage of high school teachers felt that professional development related to test preparation strategies was inadequate (30%) than did elementary or middle school teachers (23%). A greater percentage of high school teachers said there was no professional development for interpretation or use of the test results (9%) than did elementary or middle school teachers (5%). Thus, high school teachers perceived the professional development related to test preparation, interpretation and use of test results to be less adequate than did their counterparts in elementary and middle school.

Summary

The majority of teachers view the professional development related to implementation of the state-mandated testing program to be adequate. In states where the stakes are high for districts, schools, teachers, and students, more teachers view professional development as adequate than do teachers where the stakes are low for students and moderate for other groups. Conversely, greater percentages at the latter stakes levels indicate that there is no professional development focused on test preparation, interpretation, and use of test results. A smaller percentage of high school teachers also indicate that the professional development activities related to test preparation, interpretation, and use of test results is very adequate; and greater percentages of them say it is non-existent than do elementary or middle school teachers. Further, many of the differences found reflect intensity of impressions (i.e. differences in the very adequate category). Although some of the differences are small in absolute terms (4% to 5% differences), they are large in a relative sense. Higher stakes levels and lower grade levels appear to be related to greater perceived adequacy of professional development activities.

SUMMARY AND CONCLUSIONS

At least two stories emerge from these survey data. First, on several issues teachers' responses differ significantly when analyzed by the severity of the stakes attached to test results. Pressure on teachers, emphasis on test preparation, time devoted to tested content, and views on accountability are several areas where teachers' responses differed significantly by stakes level. The second story illustrates a pattern where grade level rather than stakes level reveal substantial differences in the views elementary, middle, and high school teachers hold about the effect of the state test. According to grade level, teachers' views diverge in several areas such as school climate and classroom use of test results. Further, there are some instances when stakes and grade level combined show interesting patterns in teachers' responses; in others there are no differences at all.

This section of the report is organized in the same way as the Findings section, by the major areas surveyed. These areas include (1) school climate, (2) pressure on teachers, (3) alignment of classroom practices to the state test, (4) perceived value of the state test, (5) impact of the state test on content and mode of instruction, (6) test preparation and administration, (7) unintended consequences of the state test, and (8) use of test results. Within each area, any differences among stakes levels are reported first, those among grade levels second, those among stakes and grade levels combined third, and finally overall findings are presented.

I. School Climate

Items related to school climate dealt with teacher expectations for students, student morale, how conducive the climate is to learning, student motivation, and testing pressure on students.

Stakes-level differences

Teachers' scale scores for school climate did not differ by stakes level. At the individual item level, however, there were some differences. Teachers from high-stakes states were more likely to report that students were under intense pressure to perform well on the state test and were extremely anxious about taking the state test than were teachers in M/L states. In states with high stakes for students, three-quarters or more of teachers report this intense pressure. This compares with about half of the teachers in low-stakes states. Test-related anxiety and pressure did not negatively influence teachers' expectations of student performance or their perceptions of school climate. In states where stakes are high for students, large majorities of teachers (8 in 10) reported that most of their students tried their best on the state test. Although most teachers (7 in 10) indicated that student morale was high, teachers in states with low stakes were more likely to report this than were their colleagues in high-stakes states.

Grade-level differences

Teachers' scale scores for school climate did differ by grade level. Elementary and middle school teachers were more positive (in terms of the overall scale score) about school climate than were their high school counterparts. Nonetheless, at the individual item level more elementary and middle school teachers than high school teachers reported that their students are extremely anxious and are under intense pressure because of the state test. In other words, the psychological impact was perceived to be greater at the elementary level, yet this did not seem to negatively affect the general atmosphere of the school.

Stakes level by grade level differences

There were no differences for stakes and grade level combined.

Overall

There were no overall findings of note for school climate.

II. Pressure on Teachers

Items related to pressure on teachers dealt with pressure from administrators and parents to improve test scores, pressure to limit teaching to what is tested and to change teaching strategies in ways that are not beneficial, and teachers' discontent with their profession (low morale or wanting to transfer out of tested grades).

Stakes-level differences

In general, the pressure scale shows that teachers in high-stakes states feel more pressure than those in lower-stakes states. At the individual item level, teachers did not differ by stakes level when asked about school morale or the pressure they felt from parents to raise test scores. A large majority of teachers felt that there is so much pressure for high scores on the state-mandated test that they have little time to teach anything not covered on the state test. This view was most pronounced in the H/H group. This finding supports the contention that state testing programs have the effect of narrowing the curriculum.

Teachers in high-stakes states were more likely than those in low-stakes states to report that they feel pressure from the district superintendent, and to a slightly lesser degree from their building principal, to raise test scores. While a majority of all teachers reported such pressure, it was significantly lower for teachers in low-stakes than in high-stakes states. Between 3 in 10 and 4 in 10 teachers in high-stakes states compared with 2 in 10 of their counterparts in low- stakes states reported that teachers at their school want to transfer out of the tested grades.

Grade-level differences

Generally, on the pressure scale, elementary teachers felt more pressure than did high school teachers, with middle school teachers being somewhere in between. At the individual item level, teachers did not differ by grade level when reporting on the pressure they felt from parents to raise test scores and on the morale in their school. This finding parallels the stakes-level finding. A substantial majority of teachers at each grade level indicated that state testing programs have led them to teach in ways that contradict their ideas of sound instructional practices; this view was particularly pronounced at the elementary level. This is a particularly distressing finding and one that highlights the fact that state testing programs can have unintended negative effects.

Stakes level by grade level differences

Stakes combined with grade level differences on the pressure scale result primarily from middle school teachers in the H/H and H/L states being similar to elementary teachers, and those in the H/M and M/L states being similar to their high school counterparts.

Overall

Overall, teachers, regardless of stakes or grade level, feel the greatest pressure from their superintendent.

III. Alignment of Classroom Practices with the State Test

Items related to alignment of classroom practices with the state test dealt with compatibility of the state test and curriculum, instruction, texts, and teacher-made tests.

Stakes-level differences

At the scale level, teachers in the H/H and H/L groups indicated greater alignment than did teachers in the other stakes-level groups. At the individual item level, teachers in low-stakes states more often than teachers in high-stakes states found that teaching the state standards results in better student test performance. Far more teachers in high-stakes states said their own tests reflect the format of the state test than did teachers in low-stakes states. Although the differences are not as large, a similar pattern occurs with regard to the content of teachers' tests reflecting that of the state test.

Grade-level differences

Teachers did not differ on the alignment scale by grade level. At the individual item level, elementary teachers have the most positive opinion of state curricular standards but were less positive than high school teachers about the compatibility of their instructional texts and materials with the state tests. This may be due to the fact that unlike high school teachers, who generally teach one subject, elementary teachers have to deal with several tested subjects per grade. With far more texts and materials, there is more room for disparities.

Stakes-level by grade-level differences

Teachers did not differ on the alignment scale by stakes and grade level combined.

Overall

A majority of all teachers were positive in their opinions about their state's curricular standards, and the vast majority indicated that their district's curriculum was aligned with the state test.

IV. Perceived Value of the State Test

Items related to the perceived value scale dealt with the accuracy of inferences that can be made from the test about quality of instruction, student learning, school effectiveness, and differences among various groups; the adequacy and appropriateness of media coverage of test results; and the cost/benefit ratio of the testing program.

Stakes-level differences

Teachers did not differ by stakes level on the perceived value scale. At the individual item level, teachers in high-stakes states, more so than those in low-stakes states, felt that the test brought much-needed attention to education issues. It should be noted that a minority of teachers across all stakes levels agreed with this assessment of the power of the state test to call attention to educational issues.

Grade-level differences

Teachers did not differ by grade level on the perceived-value scale. At the individual item level, elementary teachers felt to a greater degree than either middle or high school teachers that the state test measured high standards of achievement. Middle school teachers were in greater agreement with this item than were high school teachers. More elementary teachers felt that the test is not an accurate measure of what minority students know than did middle or high school teachers. Both elementary and middle school teachers felt to a greater degree than high school teachers that the test score differences from year to year reflected changes in the characteristics of students rather than changes in school effectiveness. Elementary teachers, more than middle or high school teachers, indicated that media reporting about the state test was not accurate.

Stakes-level by grade-level differences

Teachers did not differ by stakes and grade level combined on the perceived-value scale.

Overall

About three-quarters of all teachers, regardless of stakes or grade level, found that the benefits of the testing program are not worth the time and money invested. A similar proportion felt that the media coverage of issues surrounding state-mandated testing was unfair to teachers and inaccurately portrayed the quality of education and the complexity of teaching, and that score differences from year to year reflected changes in school populations more than being an indicator of school effectiveness. Across all stakes levels, 9 in 10 teachers felt that the state test was not an accurate measure of what ESL student know and can do, and 4 in 10 teachers reported that teachers in their school could raise test scores without improving learning.

V. Impact on the Content and Mode of Instruction

Items regarding the impact on classroom instruction dealt with change in the amount of time spent on a variety of activities and influence of the testing program on pedagogical practices and instructional emphasis. The items clustered into 3 scales: (1) impact on tested subject areas, (2) impact on non-core subject areas, and (3) impact on student and class activities.

Stakes-level differences

At the scale score level, more teachers in states with high stakes for students indicated that they spend more time on instruction in tested areas and less time on instruction in non-core subject areas and on other activities than did teachers in states with lesser stakes. Differences at the item level mirrored those at the scale score level. Teachers in states with high stakes for students reported increased time spent on tested areas and decreased time spent on non-tested areas (e.g. fine arts, physical education, classroom enrichment activities) to a greater degree than teachers in states with lesser stakes. In general, the influence of state testing programs on teachers' instructional practices is more closely related to the stakes for students than for schools.

Grade-level differences

At the scale score level, elementary teachers reported that they had increased the amount of time spent on instructional areas and decreased time spent on instruction in non-core subject areas and on other activities to a greater degree than high school teachers. Middle school teachers also indicated that they had increased time spent on instructional areas more than high school teachers did. The impact of testing programs on classroom instruction is generally stronger in elementary and middle schools than in high schools.

Stakes-level by grade-level differences

There were no stakes by grade level differences.

Overall

Across all types of testing programs, teachers reported increased time spent on subject areas that are tested and less time on areas not tested. They also reported that testing has influenced the amount of time spent using a variety of instructional methods such as whole-group instruction, individual-seat work, cooperative learning, and using problems similar to those on the test.

VI. Test Preparation

Teachers responded to a series of items related to preparing their students for the state-mandated test (e.g. test preparation methods used, amount of time spent on test preparation).

Stakes-level differences

Teachers in states with high-stakes tests are much more apt than their counterparts in other states to engage in test preparation earlier in the school year; spend more time on such initiatives; target special groups of students for more intense preparation; use materials that more closely resemble the test; use commercially or state-developed test-specific preparation materials; use released items from the state test; and use more motivational tactics.

Teachers in high-stakes states report spending significantly more time on test preparation than their counterparts in states where the stakes are not high. Teachers in high-stakes situations were more apt than their colleagues in low-stakes situations to report that they focused test preparation on students who were either on the border of passing or moving to the next performance level.

Grade-level differences

Elementary teachers in high-stakes situations were more likely to report spending more time in test preparation than their high school counterparts. Further, elementary teachers were more likely to report engaging in test preparation throughout the year than were middle or high school teachers.

Stakes-level by grade-level differences

Elementary teachers in states with high stakes for schools and students were twice as likely as teachers in the low-stakes states to report that their test preparation content was very similar to the content of the state test. When asked whether summer school should be required or recommended as a motivational strategy, roughly half of elementary and middle school teachers and a third of secondary teachers in the H/H states responded affirmatively. Fewer than 1 in 10 teachers across all grade levels in the low-stakes states responded "yes." Retention in grade as a motivational strategy was selected by a quarter of elementary teachers, a third of middle school teachers, and 1 in 5 high school teachers in H/H states, while the percentages in the M/L states never reached 5% at any grade level.

Experience with high-stakes tests, dating back to the 19th century and earlier, indicates that there are real dangers associated with test preparation practices. The data from this survey showing the strong relationship between the stakes associated with the test and the use of various test preparation practices are a cautionary tale that these dangers are a real possibility. Certain kinds of test preparation tactics can reduce teaching to just that — test preparation — at the expense of other subject areas, or within a subject area at the expense of material not covered by the test.

VII. Unintended Consequences of the State Test

Stakes-level differences

One-third of teachers in H/H states compared with 20% of those in M/L states said their school does not use computers when teaching writing because the state-mandated test is handwritten. Roughly one-fourth of teachers in states with high stakes for both schools and students, and one-tenth in the other high-stakes states, agreed that the test has caused retention in grades. This contrasts with only 3% of teachers in low stakes states agreeing with this statement. As for dropouts, 25% of teachers in states with high stakes for students compared with 10% of all other teachers state that the testing caused many students to drop out of high school.

Grade-level differences

There were no grade-level differences of note for the unintended-consequences items.

Stakes-level by grade-level differences

When presented with the statement that teachers in their school do not use computers when teaching writing because of the format of the state-mandated test, about one-third of middle school teachers in the H/H states agreed, as compared with 15% of their counterparts in low-stakes states. A greater percentage of teachers in states with high stakes for students agreed that the test causes students to drop out of high school. In states where the stakes are lower for students, the percentage of teachers who agreed that the test causes students to drop out decreased as grade level increased.

Overall

A majority of teachers, across the states and the stakes levels, disagreed with all of the four unintended consequences described in this section – teachers not using computers to teach writing because the state writing test is handwritten, the district forbidding the use of computers in writing instruction, the test causing many students to drop out of high school, and the test having led many students to be retained in grade.

VIII. Use of Test Results

Teachers' views on the use of test results fell into the following four categories: (1) district-level use of state test results, (2) classroom use of test results, (3) the reporting of test results, and (4) professional development and resources. Results for each of these four areas will be presented in turn; within each area, results are given for stakes level, grade level, stakes and grade levels combined, and overall.

1. Views on District-Level Use of State Test Results

Items for this area dealt with the use of state test results for three accountability purposes: school, student, and teacher/administrator accountability.

Stakes-level differences

Teachers in H/H states viewed the use of state tests for school, student, and teacher/administrator accountability as slightly less inappropriate than did teachers in other states. Further, they felt that the use of test results for student accountability was the most appropriate of the three (with a score between moderately appropriate and moderately inappropriate, a neutral view), and their use for teacher/administrator accountability was the least appropriate (having a score between moderately and very inappropriate). Although teachers in H/H states viewed the use of test results for accountability somewhat more favorably (or at least less unfavorably) than their counterparts in other states, they still fell in the neutral to unfavorable range. This more favorable view could be a result of teachers being more comfortable with tests being so used or simply being resigned to these uses. Many more teachers in H/H states said that their students' test results influence their teaching on a daily basis (25%) than did teachers in the states with lower stakes (10%). The smallest percentage of teachers who reported that the test results influence their teaching a few times a year are teachers in H/H states (10%), and the largest percentage of those who indicated that the results never influence their teaching are in low-stakes situations (15%).

Grade-level differences

High school teachers more often than elementary or middle school teachers, not surprisingly, reported that test results were used in their district to make decisions about graduation. Generally, there seemed to be less awareness at the high school level than in elementary or middle schools about how test results are used, especially how they used at the lower grade levels. This pattern may be due to the timing of decisions about placement in special education or grouping by ability, which are generally made before high school and are simply carried forward independently of state test results. This explanation, however, makes less sense for other uses (e.g. ranking schools publicly or holding schools accountable), where the district level use should be the same across all three grade levels.

Stakes-level by grade-level differences

There were no stakes by grade level differences for teachers' views on district-level use of test results.

Overall

Teachers, on average across all the states, were neutral regarding the use of state test results for student accountability. Their use for school accountability was seen on average to be moderately inappropriate, while for teacher/administrator accountability it was viewed as moderately to very inappropriate. When asked about actual uses of state tests in their districts, teachers most frequently cited use for accountability of schools and districts, ranking schools, and remediating students. Most other uses of test results were cited by less than 30% of all teachers and many by less than 10%.

2. Views on Classroom Use of State Test Results

Items for this area dealt with the influence of school- and student-level test results on teaching.

Stakes-level differences

Teachers were asked how often school-level and student-level results on the state test influenced their teaching. Significantly more teachers in states with high stakes for schools and students (40%) than in low-stakes states (10%) reported that their school's results influence their teaching on a daily basis. Conversely, a greater percentage of teachers in low-stakes states (25%) indicated that these results influence their teaching only a few times a year than teachers in states with high stakes for schools and students (roughly 10%).

Teachers in H/H states tend to use state test results for classroom decisions to a greater extent than do teachers in low-stakes situations. Teachers in states with high stakes for schools and students use the results the most of any group to plan instruction (60%) and to select instructional materials (50%); teachers in low-stakes states use them the least for these two activities (40% and 30% respectively). Teachers in states with high stakes for schools and students report using the results significantly more frequently to give feedback to students than do their counterparts in low-stakes situations. Teachers in H/H states also report using the results more often than other teachers to evaluate student progress; to group students within their class; and to determine student grades. It should be noted that the latter two uses were chosen by a small percentage of all teachers regardless of stakes level.

Grade-level differences

State-mandated test results influenced elementary teachers' instruction with much greater frequency than was the case for high school teachers. This may occur because the tests now focus elementary instruction on the standards tested, giving teachers who must teach a variety of subjects much greater direction on what should be taught. These findings may also indicate that the elementary curriculum is being narrowed or shaped to a greater degree by state-mandated tests than is the case at the high school level. Conversely, high school teachers' instruction may be least influenced by the state tests, because these teachers have always taught a specific subject area (e.g. math or history), and the test is measuring, for the most part, the content they were already teaching. Middle school teachers fall somewhere between elementary and high school teachers in terms of subject matter specialization, and therefore the influence of the state test results on their instruction is somewhere between that for the other two groups, although generally closer to the elementary level.

More elementary teachers reported using the results of the state-mandated test to aid in decisions about instruction, assess their own teaching effectiveness, provide feedback to parents, evaluate students, and group students in their class than did high school teachers. In general, high school teachers are least likely to use state test results.

Stakes-level by grade-level differences

Teachers' views on classroom use of the state test results did not differ by stakes and grade levels combined.

Overall

The test results are used by about one-third of all teachers to assess their teaching effectiveness and give feedback to parents or students. Between 40% and 50% of all teachers reported using the results to plan instruction or curriculum or to select instructional materials. Clearly, the stakes attached to the results of the state-mandated test affect the extent to which teachers use them for various instructional and feedback activities. When the stakes are high for students and teachers, teachers use the results to the greatest extent; when they are low, they tend to use them less often. For 7 of the 8 activities listed, fewer than half of the teachers – regardless of stakes level – indicated that they use the test results to inform their practice — the lone exception being that a majority of all teachers reported using results to plan instruction. Further, very small proportions (less than 10% overall) use the results for student-specific decisions (i.e. grouping students within the class or determining student grades).

3. Views on the Reporting of Test Results

Items for this section dealt with the various test-result reports that teachers receive: individual student reports, school reports, and district-level reports.

Stakes-level differences

A majority of all teachers either agree or strongly agree that the individual student, school, and district reports are easy to interpret and provide useful information. A significantly larger proportion of teachers (though still small at 10%) in the states with low stakes were unfamiliar with the school and district reports than were teachers in any of the three high-stakes groups.

Grade-level differences

High school teachers are the least familiar with the various reports. Between 10% and 20% report they have never seen them. Significantly fewer high school teachers than elementary or middle school teachers agreed that the reports provide useful information. Elementary teachers have the greatest familiarity with the school reports, less than 10% indicating that they had never seen them.

Stakes-level by grade-level differences

There were no stakes combined with grade level differences on views on the reporting of results.

Overall

There were no overall findings of note for the reporting of test results.

Professional Development and Resources

Items for this section dealt with the adequacy of professional development around the state testing program and the availability of someone in the school to deal with questions about the program.

Stakes-level differences

The vast majority of all teachers (80%) indicated that they do have someone to turn to at their school to obtain accurate information about the state-mandated testing program. The sole difference occurred between teachers in states with high stakes for students and schools and those in states with low stakes (80% vs. 70%). A greater percentage of teachers in states where the stakes are high viewed the professional development as adequate than of teachers where the stakes are low. Conversely, greater proportions of teachers in low-stakes situations indicated that there is no professional development related to test preparation, interpretation, and use of test results.

Grade-level differences

A significantly smaller percentage of high school teachers also indicated that the professional development activities around test preparation, interpretation, and use of test results are adequate than did elementary or middle school teachers.

Stakes-level by grade-level differences

There were no stakes combined with grade level differences on views on the reporting of results.

Overall

The majority of all teachers view the professional development related to areas concerning implementation of the state-mandated testing program to be adequate.

Conclusions

As indicated at the beginning of this section, we found differences attributable to stakes level, grade level, and the interaction of these two levels. For some items or scales, there were no differences among these levels; these findings were also of interest.

For the most part, teachers in states with high stakes for both students and teachers (or schools and districts), i.e. H/H teachers, held views about the effect of state testing programs that differed from those of teachers in states where the stakes were low. The differences were in the expected direction: teachers in high-stakes situations, particularly in H/H states, reported feeling more pressure to have their students do well on the test, to align their instruction with the test, to engage in more test preparation, and so forth. In many instances, results from teachers in states where the stakes were low for students but high for schools (H/L) were very similar to those for teachers in H/H states.

Elementary teachers often indicated that they are greatly affected by the statewide testing program. For example, they reported increased time spent on instruction in tested areas, less time spent on instruction in non-tested areas, more time spent on test preparation, greater impact on their instructional practices, and so on than did secondary teachers.

The findings in this report need to be examined by policymakers and educators in their own state to determine whether the effects of the state test, as reported here by teachers, are desired. To the extent that undesired effects are occurring, the testing program should be modified so as to minimize them. Only by listening to what teachers tell us is happening as a result of these state testing programs can we be confident that they are having the intended effect. Teachers are on the front line every day. Their voice on this issue must be heard; their opinions must enter into the formation of sound testing policy. Although some states do involve teachers in the formulation of their testing program, others do not. Even in those states where teachers are involved, the number of teachers is small. We hope the findings presented here give voice to a broader cross-section of teachers than has heretofore been available on issues related to state-wide testing programs, and spur more teacher input in the future.

END NOTES

1 In another study that surveyed teachers, Hoffman, Assaf, and Paris (2001) obtained a mail survey return rate of 27%.

2 This item loaded similarly on both the school climate and pressure on teachers scale. However, item 13 was included in the pressure scale, since all of the items on this scale focused specifically on teachers and were either directly or indirectly associated with feelings of pressure.

3 Reverse coding was not necessary when computing scale scores for the alignment scale, since the items were generally neutral or positive statements.

REFERENCES

Barksdale-Ladd, M. & Thomas, K. (2000). What's at stake in high stakes testing? *Journal of Teacher Education 51*, 384-97.

Center on Education Policy (2002). *State high school exit exams: A baseline report.* Washington, DC: Center on Education Policy.

Clarke, M., Abrams, L., & Madaus, G. (2001). The effects and implications of high-stakes achievement tests for adolescents. In T. Urdan & F. Pajares (Eds.). *Adolescence and education: Vol. 1. General issues in the education of adolescents* (pp.201-229). Greenwich, CT: Information Age Publishing.

Clarke, M., Haney, W., & Madaus, G. (2000). *High stakes testing and high school completion.* Chestnut Hill, MA: National Board on Educational Testing and Public Policy.

Corbett, H., & Wilson, B. (1991). *Testing, reform, and rebellion.* Norwood, NJ: Ablex.

Dillman, D. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley & Sons.

Fassold, M (1996). *Adverse racial impact of the Texas Assement of Academic Skills.* San Antonio, TX: Mexican American Legal Defense and Education Fund.

Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113.

Goertz, M. (2000, April). Local accountability: *The role of the district and school in monitoring policy, practice, and achievement.* Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.

Greaney, V., & Kellaghan, T. (1996). Equity and the integrity of public examinations in developing countries. In H. Goldstein & T. Lewis (Eds.). *Assessment: Problems, developments and statistical issues.* Chichester, England: Wiley.

Haladyna, T., Nolen, S., & Hass, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20*(5), 2-7.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives,* 8(41). Retrieved April 13, 2001 from http://epaa.asu.edu/epaa/v8n41

Herman, J., & Golan, S. (n.d.). *Effects of standardized testing on teachers and learning: Another look.* (CSE Technical Report 334). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Heubert, J. & Hauser, R. Eds. (1999). High-stakes: *Testing for tracking, promotion and graduation.* Washington, DC: National Academy Press.

Hoffman, J., Assaf, L., & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher, 54*(5), 482-494.

Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis, 23* (2), 99-121.

Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*(3), 199-203.

Kellaghan, T. & Greaney, V. (1992). *Using examinations to improve education: A study in fourteen African countries.* Washington, DC: World Bank.

Kellaghan, T., Madaus, G., & Airasian, P. (1980). *The effects of standardized testing.* Educational Research Centre: St. Patrick's College, Dublin, Ireland and Boston College: Chestnut Hill, MA.

Kellaghan, T., Madaus, G., & Raczek, A. (1996). *The use of external examinations to improve student motivation.* Washington, DC: American Educational Research Association.

Klein, S., Hamilton, L., McCaffrey, D., & Stecher, B. (2000). *What do test scores in Texas tell us?* Santa Monica, CA: RAND.

Koretz, D. & Barron, S. (1998). *The validity gains on the Kentucky Instructional Results Information System (KIRIS).* (MR-1014-EDU). Santa Monica, CA: RAND.

Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996a) *Perceived effects of the Kentucky Instructional Results Information System (KIRIS).* (MR-792-PCT/FF). Santa Monica, CA: RAND.

Koretz, D., Linn, R., Dunbar, S., & Shepard, L. (1991, April). *Effects of high-stakes testing on achievement: Preliminary findings about generalization across tests.* Paper presented at the annual meeting of the American Education Research Association and the National Council of Measurement in Education, Chicago.

Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996b). *Final report: Perceived effects of the Maryland school performance assessment program.* (CSE Technical Report 409). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). *The evolution of a portfolio program: The impact and quality of the Vermont program in its second year (1992-93).* (CSE Technical Report 385). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Kreitzer, A., Madaus, G., & Haney, W. (1989). Competency testing and dropouts. In Weis, L., Farrar, E. & Petrie, H. (Eds.). *Dropouts from school: Issues dilemmas and solutions.* (pp. 129-152). Albany State: State University of New York Press.

Lane, S, Parke, C. & Stone, C. (1998, April). Consequences of the Maryland school performance assessment program. Paper presented at the annual meeting of the National Council of Measurement in Education, San Diego, CA.

Linn, R. (2000). Assessments and accountability. *Educational Researcher, 29* (2), 4-16.

Madaus, G. (1988). The influence of testing on the curriculum. In Tanner, L. (Ed.). *Critical issues in curriculum.* (pp. 83-121). Chicago, IL: University of Chicago Press.

Madaus, G. (1991, January). *The effects of important tests on students: Implications for a national examination or system of examinations.* Paper prepared for the American Educational Research Association Invitational Conference on Accountability as a State Reform Instrument: Impact on Teaching, Learning, Minority Issues and Incentives for Improvement. Washington, D.C.

Madaus, G., & Greaney, V. (1985). The Irish experience in competency testing: Implications for American education. *American Journal of Education,* 93 (2), 268-294.

Madaus, G., & Kellaghan, T. (1992). A national testing system: Issues for the social studies community. *Social Education,* 56 (2), 89-91.

Madaus, G., West, M., Harmon, M., Lomax, R., & Viator, K. (1992). *The influence of testing on teaching math and science in grades 4-12.* Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

McMillan, J., Myran, S., Workman, D. (1999, April). *The impact of mandated statewide testing on teachers' classroom assessment and instructional practices.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

McNeil, L.(2000). *Contradictions of school reform: Educational costs of standardized testing.* New York. Routledge.

Mehrens, W. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives,* 6 (13). Retrieved August 14, 2000 from http://epaa.asu.edu/epaa/v6n13.html

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of Children, 5* (2), 113-27.

Noble, A., & Smith, M. (1994). *Old and new beliefs about measurement-driven reform: "The more things change, the more they stay the same."* (CSE Technical Report 373). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Quality Counts 2002 (2002, January 10). *Education Week, 21* (17).

Reardon, S. (1996, April). *Eighth grade minimum competency testing and early high school drop out patterns.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Shepard, L. (1990). Inflating test score gains: Is the problem old norms or teaching the test. *Educational Measurement: Issues and Practice, 9* (3), 15-22.

Shepard, L. (2001). The role of assessment in teaching and learning. In V. Richardson (Ed.), *Handbook on research on teaching* (4th ed.). Washington, DC: American Educational Research Association.

Shore, A., Pedulla, J., & Clarke, M. (2001). *The building blocks of state testing programs.* Chestnut Hill, MA: National Board on Educational Testing and Public Policy.

Smith, M. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20* (5), 8-11.

Smith, M., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1991). *The role of testing in elementary schools.* (CSE Technical Report 321). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Smith, M., Nobel, A., Heinecke, W., Seck, M., Parish, C., Cabay, M. et al. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona student assessment program (ASAP): Equity and teacher capacity building.* (CSE Technical Report 425). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Smith, M. & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice, 10* (4), 7-11.

Stecher, B., Barron, S., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classrooms.* (CSE Technical Report 525). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Stecher, B., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effect of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing.* (CSE Technical Report 482). National Center for Research on Evaluation, Standards and Student Testing.

Urdan, T. & Paris, S. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy, 8*(2), 137-156.

U.S. Department of Education, National Center for Education Statistics. (2002). *Digest of education statistics 2001.* Washington, DC.

Wolf, S., Borko, H., McIver, M., & Elliot, R. (1999). *No excuses: School reform efforts in exemplary schools of Kentucky.* (CSE Technical Report 514) National Center for Research on Evaluation, Standards and Student Testing.

Appendix A (Teacher Survey) is not included in this online version due its large memory size. It can be downloaded at http://www.bc.edu/research/nbetpp/reports.html.

APPENDIX B
STATE TESTING PROGRAM CLASSIFICATION GRID

CONSEQUENCES FOR STUDENTS

CONSEQUENCES FOR TEACHERS, SCHOOLS, AND DISTRICTS	High	Moderate	Low
High	Alabama California* Delaware* Florida Georgia* Indiana* Louisiana Maryland* Massachusetts* Mississippi* Nevada New Jersey New Mexico New York North Carolina South Carolina Tennessee Texas Virginia*	Arkansas Connecticut Illinois Michigan Pennsylvania West Virginia	Colorado* Kansas Kentucky Missouri Oklahoma* Rhode Island Vermont*
Moderate	Arizona* Alaska* Ohio Minnesota Washington* Wisconsin*	Oregon	Hawaii Maine Montana Nebraska New Hampshire North Dakota South Dakota Utah* Wyoming
Low	Idaho*		Iowa

**Indicates that the program was not fully in place at the time of this study.*

APPENDIX C

Sample Stratifications and Final Sampling Frame

In addition to the basic sampling frame based on stakes levels (Table 1), teachers were further randomly selected according to the type of school in which they taught (elementary, middle and high), subject area (high school teachers), and geographic setting of the school (urban or non-urban area). The following tables illustrate the various stratifications of the sample.

Table C1. Sampling Stratification by School Type

Stakes Level	Elementary School	Middle School	High School	Total
H/H	550	550	1,100	2,200
H/M	550	550	1,100	2,200
H/L	550	550	1,100	2,200
M/H	550	550	1,100	2,200
M/L	550	550	1,100	2,200
Massachusetts	250	250	500	1,000
School Level Totals	3,000	3,000	6,000	12,000

Table C2 illustrates a further detailing of the sample incorporating the course content areas taught by teachers at the high school level.

Table C2. Sampling Stratification by School Type and Subject Area

Stakes Level	Elementary School	Middle School	High School					Total
			English	Math	Science	Soc. Stud.	Special Ed.	
H/H	550	550	220	220	220	220	220	2,200
H/M	550	550	220	220	220	220	220	2,200
H/L	550	550	220	220	220	220	220	2,200
M/H	550	550	220	220	220	220	220	2,200
M/L	550	550	220	220	220	220	220	2,200
MA	250	250	100	100	100	100	100	1,000
School Level Totals	3,000	3,000	1,200	1,200	1,200	1,200	1,200	12,000

Table C3 depicts the 84 segments of the final sampling frame that included all of the stratifying variables (stakes levels, school type, and subject area) proportionally across urban and non-urban areas.

Table C3. Final Sampling Frame

Stakes Level	Elementary School		Middle School		High School										Total
					English		Math		Science		Soc. Stud.		Special Ed.		
	U	NU	U	NU	U	NU	U	NU	U	NU	U	NU	U	NU	
H/H	194	356	194	356	78	142	78	142	78	142	78	142	78	142	2,200
H/M	146	404	146	404	58	162	58	162	58	162	58	162	58	162	2,200
H/L	131	419	131	419	52	168	52	168	52	168	52	168	52	168	2,200
M/H	180	370	180	370	72	148	72	148	72	148	72	148	72	148	2,200
M/L	121	429	121	429	48	172	48	172	48	172	48	172	48	172	2,200
MA	70	180	70	180	28	72	28	72	28	72	28	72	28	72	1,000
Subtotals	842	2,158	842	2,158	336	864	336	864	336	864	336	864	336	864	12,000
School Level Totals	3,000		3,000		1,200		1,200		1,200		1,200		1,200		12,000

APPENDIX D

Table D1. Characteristics of Survey Respondents[1]

Respondent Characteristics		N	% of Respondents	% of Population
Gender[2]	Male	764	18	26
	Female	3,396	81	74
Age[2]	20-30	520	12	11
	31-40	816	20	
	41-50	1,325	32	67 (are 40 or older)
	51-60	1,356	33	
	60+	130	3	
Race/Ethnicity[3]	African American	298	7	7
	American Indian/ Alaskan Native	57	1	1
	White	3,621	86	91
	Asian/Pacific Islander	39	1	1
	Hispanic	199	5	4
Grade Level[2]	Elementary School	2,448	58	60
	Middle School	836	20	40 (secondary)
	High School	911	22	
Content Area of High School Teachers[3,4]	English	368	40	24
	Math	214	23	17
	Science	139	15	13
	Social Studies	165	18	13
	Special Education	149	16	2
Teaching Experience (years)[3]	1	71	2	17 (5 years or less)
	2-3	284	7	
	4-8	723	17	Average is 16 years
	9-12	508	12	
	13-20	898	22	
	20+	1,679	40	46
School Location[5]	Urban	1,108	26	32
	Suburban	1,782	43	38
	Rural	1,304	31	30
Testing Stakes for Teachers, Schools, Districts/Stakes for Students[5]	High/High	2,549	61	60
	High/Moderate	642	15	15
	High/Low	355	9	9
	Moderate/High	471	11	12
	Moderate/Low	180	4	4

1. Numbers are weighted using estimates of the national population.

2. Population estimates based on NEA Rankings Estimates: Rankings of the States 2000 and Estimates of School Statistics 2001

3. Population estimates based on NEA Status of the American Public School Teacher, 1995-96: Highlights and Digest of Education Statistics, 2000.

4. Total percent of respondents exceeds 100 because some high school teachers reported teaching more than one content area.

5. Market Data Retrieval population estimates, fall 2000

APPENDIX E

Table E1. School Climate Scale Summary[1]

School Climate-Related Survey Items	Factor Loadings
My school has an atmosphere conducive to learning.	-.695
Teachers have high expectations for the in-class academic performance of students in my school.	-.681
The majority of students try their best on the state-mandated test.	-.560
Student morale is high in my school.	-.466
Teachers have high expectations for the performance of all students on the state-mandated test.	-.442
Many students are extremely anxious about taking the state-mandated test.	-.441
Students are under intense pressure to perform well on the state-mandated test.	-.471
Many students in my school cheat on the state-mandated test.	.415

[1] *The Cronbach alpha reliability for the scale was .64.*

Table E2. ANOVA Results for Stakes Level and School Type on the School Climate Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	1.50	4	.38	3.67	.006
School Type	6.69	2	3.47	32.66	.000
Stakes by School Type	1.71	8	.21	2.09	.033
Error	425.49	4152	.102		
Total	33682.17	4167			

Table E3. Pressure Scale Summary[1]

Pressure-Related Survey Items	Factor Loadings
Teachers feel pressure from the building principal to raise scores on the state-mandated test.	.716
Teachers feel pressure from the district superintendent to raise scores on the state-mandated test.	.617
Administrators in my school believe students' state-mandated test scores reflect the quality of teachers instruction	.592
The state-mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good educational practice.	.589
There is so much pressure for high scores on the state-mandated test teachers have little time to teach anything not on the test.	.578
Teacher morale is high in my school.	-.557
Teachers in my school want to transfer out of the grades where the state-mandated test is administered.	.546
Teachers feel pressure from parents to raise scores on the s tate-mandated test.	.218

1. The Cronbach alpha reliability for the scale was .75.

Table E4. ANOVA Results for Stakes Level and School Type on the Pressure Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	30.56	4	7.64	37.85	.000
School Type	21.52	2	10.76	53.29	.000
Stakes by School Type	7.57	8	.96	4.739	.000
Error	821.05	4066	.20		
Total	33432.73	4080			

Alignment-Related Survey Items	Factor Loadings
My district's curriculum is aligned with the state-mandated testing program.	.722
The state-mandated test is compatible with my daily instruction.	.695
The state-mandated test is based on a curriculum framework that all teachers should follow.	.616
My tests have the same content as the state test.	.608
The instructional texts and materials that the district requires me to use are compatible with the state-mandated test.	.598
My tests are in the same format as state test.	.573

Table E5. Alignment Scale Summary[1]

[1] *The Cronbach alpha reliability for the scale was .73.*

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	16.11	4	4.03	19.42	.000
School Type	0.58	2	0.29	1.41	.246
Stakes by School Type	0.65	8	8.150E-02	0.39	.925
Error	841.42	4058	0.21		
Total	28383.01	4073			

Table E6. ANOVA Results for Stakes Level and School Type on the Alignment Scale

Table E7. Perceived-Value Scale Summary[1]

Value-Related Survey Items	Factor Loadings
Overall, the benefits of the state-mandated testing program are worth the investment of time and money.	.698
Media coverage of state-mandated test results accurately reflects the quality of education in my state.	.573
Scores on the state-mandated test results accurately reflect the quality of education students have received.	.566
The state-mandated test has brought much needed attention to education issues in my district.	.542
The state-mandated test is as accurate a measure of student achievement as a teacher's judgment.	.539
The state-mandated test motivates previously unmotivated students to learn.	.530
The state-mandated test measures high standards of achievement.	.516
The state-mandated testing program is just another fad.	-.461
Media coverage of state-mandated testing issues has been unfair to teachers.	-.430
Media coverage of state-mandated testing issues adequately reflects the complexity of teaching.	.420
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	-.375
The state-mandated test is not an accurate measure of what students who are acquiring English as a second language know and can do.	-.308
Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness.	-.269

1. The Cronbach alpha reliability for the scale was .79.

Table E8. ANOVA Results for Stakes Level and School Type on the Perceived-Value Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	2.13	4	.53	3.88	.004
School Type	.50	2	.25	1.84	.159
Stakes by School Type	2.64	8	.33	2.41	.014
Error	562.43	4106	.14		
Total	16946.89	4121			

Table E9. Tested Areas, Non-Core Content, Classroom Activities Scales Summary

Item 62: In what ways, if any, has the amount of time spent on each of the following activities changed in your school in order to prepare students for the state-mandated testing program?	Scales and Factor Loadings		
	Test-Content Areas	Non-Core Content Areas	Activities
Instruction in tested areas	-.710		
Instruction in tested areas with high stakes attached (e.g., promotion, graduation, teacher rewards)	-.651		
Parental contact	-.573		
Instruction in areas not covered by the state-mandated test	-.536		
Instruction in physical education		.808	
Instruction in foreign language		.803	
Instruction in industrial/vocational education		.777	
Instruction in fine arts		.759	
Enrichment school assemblies (e.g., professional choral group performances)			.782
Class trips (e.g., circus, amusement park)			.779
Field trips (e.g., museum tour, hospital tour)			.767
Student choice time (e.g., games, computer work)			.756
Organized play (e.g., games with other classes)			.752
Classroom enrichment activities (e.g., guest speakers)			.742
Student performance (e.g., class plays)			.742
Administrative school assemblies (e.g., awards ceremonies)			.713
Student free time (e.g., recess, lunch)			.511
Scale Reliability (Cronbach's alpha)	.57	.83	.91

Table E10. ANOVA Results for Stakes Level on the Tested Areas Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	67.97	4	16.99	64.45	.000
School Type	12.89	2	6.45	24.45	.000
Stakes by School Type	3.17	8	.40	1.5	.150
Error	947.61	3594	.26		
Total	50071.00	3609			

Table E11. ANOVA Results for Stakes Level and School Type on the Non-Core Content Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	33.43	4	8.36	28.44	.000
School Type	4.67	2	2.33	7.94	.000
Stakes by School Type	5.83	8	.73	2.48	.011
Error	967.37	3292	.29		
Total	27222.94	3307			

Table E12. ANOVA Results for Stakes Level and School Type on the Classroom Activities Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	44.17	4	11.04	36.70	.000
School Type	5.10	2	2.55	8.47	.000
Stakes by School Type	7.25	8	.91	3.01	.002
Error	1159.62	3853	.30		
Total	27837.53	3868			

Table E13. School, Student, Teacher/ Administrator Accountability Scales Summary

Item 61: The following is a list of ways in which state-mandated test results are used for each item. Please indicate how appropriate you feel the specific use is.	School Accountability Scale	Student Accountability Scale	Teacher/Admin Accountability Scale
Evaluate charters schools	.840		
Evaluate voucher programs	.804		
Hold the district accountable	.850		
Hold schools accountable	.842		
Award school accreditation	.744		
Place schools in receivership	.647		
Rank schools publicly	.631		
Place students in special education		.755	
Place students in gifted programs		.695	
Promote/retain students in grade		.756	
Remediate students		.684	
Group students by ability in grade		.651	
Graduate students from high school		.685	
Award teachers/admin. financial bonuses			.858
Reward schools financially			.838
Evaluate teacher/admin. performance			.789
Fire faculty/staff			.708
Scale Reliability (Cronbach's alpha)	.89	.80	.84

Table E14. ANOVA Results for Stakes level and School Type on the School Accountability Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	33.59	4	8.40	19.36	.000
School Type	2.48	2	1.24	2.86	.058
Stakes by School Type	2.85	8	.36	.821	.584
Error	1752.50	4041	.43		
Total	15388.94	4055			

Table E15. ANOVA Results for Stakes Level and School Type on the Student Accountability Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	51.05	4	12.76	30.90	.000
School Type	.74	2	.369	.89	.409
Stakes by School Type	3.79	8	.474	1.15	.328
Error	1718.30	4161	.413		
Total	24265.13	4175			

Table E16. ANOVA results for Stakes Level and School Type on the Teacher/ Administrator Accountability Scale

Sources of Variation	SS	df	MS	F-ratio	Signif.
Stake Level	41.82	4	10.46	35.94	.000
School Type	1.99	2	.99	3.42	.033
Stakes by School Type	1.29	8	.16	.55	.817
Error	1202.94	4135	.29		
Total	9090.625	4149			

The National Board on Educational Testing and Public Policy

NBETPP

About the National Board on Educational Testing and Public Policy

Created as an independent monitoring system for assessment in America, the National Board on Educational Testing and Public Policy is located in the Carolyn A. and Peter S. Lynch School of Education at Boston College. The National Board provides research-based test information for policy decision making, with special attention to groups historically underserved by the educational systems of our country. Specifically, the National Board

- Monitors testing programs, policies, and products
- Evaluates the benefits and costs of testing programs in operation
- Assesses the extent to which professional standards for test development and use are met in practice

This National Board publication is supported by a grant from The Atlantic Philanthropies Foundation.

The National Board on Educational Testing and Public Policy
Lynch School of Education, Boston College
Chestnut Hill, MA 02467

Telephone: (617)552-4521
Fax: (617)552-8419
Email: nbetpp@bc.edu

BOSTON COLLEGE

Visit our website at **www.bc.edu/nbetpp** for more articles, the latest educational news, and information on NBETPP.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (1/2003)